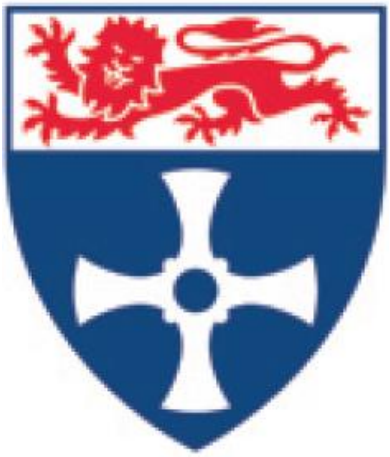


Detailed analysis of sequence requirements within 2A translational recoding peptides

Fu Yan

Institute for Cell and Molecular Biosciences

Newcastle university

Thesis submitted to Newcastle University in candidature for the degree of Doctor of Philosophy

April 2013

Declaration

I hereby declare that I alone have composed the thesis and that, except where stated, the work presented herein is my own.

Fu Yan

Abstract

2A sequences are short 20 ~30 amino acid peptides initially characterised from foot and mouth disease virus (FMDV), but also encoded in a range of Picornaviruses and other viruses as well as non-LTR retrotransposons in a broad range of organisms. They direct a translational recoding event in which ribosomes that have reached the final codon of the 2A sequence pause and terminate translation in the absence of a stop codon and then restart translation. The effect is to separate a single ORF into 2 polypeptides, between the final 2 amino acids (glycine and proline) of 2A, 'skipping' a peptide bond. 2A sequences comprise 2 parts, a conserved GDV/IEXNPGP C-terminal motif, and a highly variable, but necessary, N-terminal portion. In the work described in this thesis, extensive mutagenesis studies were carried out across both the conserved and non-conserved portions of the FMDV 2A peptide, including alanine, glycine and proline scanning mutagenesis, site-directed mutagenesis screens. These data indicate that while the amino acid present at the first 2-3 positions of the peptide is not critical for activity, the residue at most other positions is important, including several positions that vary considerably between different 2A peptides. Intriguingly, results of alanine, glycine and proline scanning mutagenesis were very similar, suggesting that features of side chains, and not just the secondary structure propensity of the peptide, plays a critical role. Despite their great variability, each 2A peptide must provide a set of interactions between side chains within the peptide and with the ribosomal exit tunnel to allow the 2A reaction to proceed. Finally, the work also verified a previous finding that insertion of a stop codon in place of the final proline codon of 2A leads ribosomes to stall, unable to terminate translation. Efforts were made to establish a system for purification of such stalled ribosome-nascent chain complexes towards the eventual aim of structural characterisation. The data strongly support a model in which the interaction of the 2A peptide with the ribosome distorts ribosomal conformation.

Acknowledgements

First of all, I thank my supervisor Dr. Jeremy D Brown. He is a true scientist and a great man. His passion encourages me in every moment of my project, his patience and kindness is beyond my imagination, and his preciseness and prudence to science will benefit me forever. My special thanks goes to Vicki and Pamila, in the early stage of my project on 2A, they open the door of this wonderful subject to me; and Hatem, for he generously sharing his great ideas to me; and all the members of Brown lab, without the encouragements and friendships sourced from **THE LAB**, this job would not be possible. I also thank all members of Watkins lab and all members of yeast lab (M3023), who made the lab an excellent working environment.

I also wish to thank the faculty of the medical school. I've got numerous supports from college. Especially I wish to thank Dr. Tim Cheek, who, regrettably I have to say, I troubled the most. This work also obtained many helps outside Newcastle, Prof. Martin Ryan of St. Andrews shared the excellent vector pLH135/136 to us, which has became the housekeeping plasmids for this study. Thank you!

Finally, I wish to express my gratitude to my parents, not only for raising and support me in last 28 years, but also for their courage in face of formidable disease, which did and will forever encourage me.

M.

Fu Yan

Contents

List of Figures

List of Tables

Glossary of terms

ATP	adenosine 5’-triphosphate
Amp	ampicillin
APS	ammonium persulphate
bp	base pair
CMEF	complete EDTA-free mini protease inhibitor tablets
CFP	cyan fluorescent protein
dH_2O	distilled water
ddH_2O	double distilled water
$dNTP_S$	dideoxynucleoside triphosphate
DNA	deoxynucleic acid
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
g	gram(s)
GFP	green fluorescent protein
GTP	guanine 5’-triphosphate
H_2O	hydrogen peroxide
kb	kilo base
kDa	kilodaltons
L	litre(s)

LB	Luria-Bertani medium
LTRs	long terminal repeat retrotransposons
M	mole
µg	micro gram(s)
µl	micro litre(s)
MES	2-(4-Morpholino)ethanesulfonic acid
min	minute(s)
mg	milligram(s)
ml	millilitre(s)
MOPS	Morpholinopropanesulfonic acid
mRNA	messenger ribonucleic acid
nt	nucleotides
OD	optical density
ORF	open reading frame
PAC	puromycin N-acetyltranferase
PCI	phenol:chloroform:isoamylalcohol
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulphonyl fluoride
PTC	peptidyl transferase center
RNA	ribonucleic acid
RNAse	ribonuclease
PTC	peptidyl-transferase centre

rpm	revolution per min
sec	seconds
SDS	sodium dodecyl sulphate
TBS	tris buffered saline
TCA	tricholoroacetic acid
TEMED	tetramethylethy lenediamine
Tris	tris aminomethance
tRNA	transfer ribonucleic acid
Ub	ubiquitin
V	volt(s)
v/v	volume per unit volume
w/v	weight per unit volume
YPD	yeast peptone dextrose

Amino acid nomenclature

Nomenclature	3-letter code	1-letter code
Alanine	**Ala**	**A**
Arginine	**Arg**	**R**
Asparagine	**Asn**	**N**
Aspartic acid	**Asp**	**D**
Cysteine	**Cys**	**C**
Glutamic acid	**Glu**	**E**
Glutamine	**Gln**	**Q**
Glycine	**Gly**	**G**
Histidine	**His**	**H**
Isoleucine	**Ile**	**I**
Leucine	**Leu**	**L**
Lysine	**Lys**	**K**
Methionine	**Met**	**M**
Phenylalanine	**Phe**	**F**
Proline	**Pro**	**P**
Serine	**Ser**	**S**
Threonine	**Thr**	**T**
Tryptophan	**Trp**	**W**
Tyrosine	**Tyr**	**Y**

Valine	**Val**	**V**

Chapter 1 Introduction

The transfer of information from DNA to RNA then to protein is regarded as the central dogma of molecular biology (Crick *et al*., 1970). This procedure, however, is not a linear, simple one, instead, it is a complicated course of many events, in which great number of factors are involved. In eukaryotic cells, RNA is transcribed by three polymerases: RNA polymerase I and III, which transcribe the RNAs that will not be further decoded for proteins (non-coding RNA), such as rRNAs forming the RNA section of ribosome, or tRNA involved in translation and RNA polymerase II, which mainly transcribes the mRNA precursor that encode proteins (Dieci *et al*., 2007; Sims *et al*., 2004; Russell *et al*., 2006). The precursor mRNA is not used in translation directly, it has to be processed by several modifications before translation. The main modifications are addition of 5' 7-methyl-guanosine cap and 3' poly-A tail, which are critical for translation and also protect the mRNA from nucleases, and RNA splicing that removes the introns from the pre-mRNA (Cogoni *et al*., 2000; Helm *et al*., 2006).

Mature mRNA directs the ribosome to synthesise the peptide chain corresponding to its nucleotide sequence. Occasionally, however, the outcome of protein synthesis does not match that predicted from inspection of the mRNA sequence. Such phenomena are termed as translational recoding events. Most translational recoding events are driven by features of the mRNA, such as stem loops or pseudoknots adjacent to the recoding site. These are, though, recoding events directed by the amino acid sequence of the nascent chain. The 2A reaction, analysed in this thesis, falls within this category.

The 2A reaction is directed by a collection of short peptides (20~30 amino acids long) termed 2A peptides. During the translation, they cause a 'skip' in peptide bond formation at the C-terminus of their sequences, which results the generation of two polypeptides from one open reading frame, an extremely unusual event during

translation. The 2A reaction was initially found during the study of foot and mouse disease virus, a picornavirus that infects cloven-hoofed animals. While playing critical roles in the life-cycle and reproduction of many picornavirus as well as a number of other viruses, 2A peptides are also found in the genome of other organisms, such as cellular genes of sea urchin and non-LTR retrotransposons in a wide range of both unicellular and multicellular organisms (Heras *et al*., (2006) and M. Ryan personal communication). Interestingly, 2A peptides only promote the 2A reaction in concert with the eukaryotic translational apparatus.

In the introduction of this thesis, procedure of the 'textbook' model of protein translation are described, for offering the background information of 2A reaction; (section 1.1) followed by a summary of different types of translational recording events as well as other unconventional translational mechanisms that are relative to 2A (section 1.2). The final part (section 1.3) of this chapter summarizes the previous research on 2A reaction and 2A peptides.

1.1 Translation and the ribosome

Protein synthesis, the biogenesis of protein with mRNA as template is a sequential and precise procedure, with numerous biological macromolecules involved. The ribosome, the platform where translation occurs, is the centre of the translational apparatus.

1.1.1 Structure and function of the ribosome

The ribosome, is a large ribonucleoprotein particle, and in all species comprises two subunits. In eukaryotes, the subunits are designated 60S and 40S, and together make up 80S ribosome complex. The 40S subunit is composed of one RNA component (18S rRNA) and 33 ribosomal proteins (rprotein) whereas the 60S subunit has 3 rRNA (5S, 28S and 5.8S respectively) and 46 rproteins (Ben-Shem *et al*., 2011; Klinge *et al*., 2011; Rabl *et al*., 2010). The core structure of ribosome from all species is universally

conserved, and the rRNA sequences beyond the core structure, that are different between eukaryotic, prokaryotic and archaebacterial ribosomes, are termed expansion segments. During translation, the small subunit directs amino acyl-tRNAs to read the mRNA sequence, and the large subunit hosts tRNA and catalyzes the formation of the peptide bond. On the interface between two subunits, there are three tRNA binding sites, formed by ribosomal RNA. The tRNA-binding sites are: the aminoacyl site (A site) decoding the mRNA through the codon-anticodon base pairing, and working as the dock of the ribosome for the incoming amnoacyl tRNA; peptidy site (P site), holding the tRNA with the growing peptide chain (peptidyl-tRNA); and exit site (E site), the site from which the deacylated, uncharged tRNA leaves ribosome (Kim *et al*., 1999; Berk *et al*., 2006). The principal chemical reaction of protein synthesis, peptide bond formation, is carried out by the peptidyl transferase centre (PTC), a region of large subunit that is adjacent to the P and A sites. During translation, it catalyzes bond formation between the amino acids of aminoacyl tRNA in A site, and the nascent peptide chain of peptidyl-tRNA in P site, with assistance of several elongation factors (Figure 1.1) (Beringer *et al*., 2007; Polacek *et al*., 2005). This function, is not mediated by rprotein, but by rRNA (in eukaryotic ribosome by 28S rRNA of the large unit) (Noller *et al*., 1993; Simonovic *et al*., 2009).

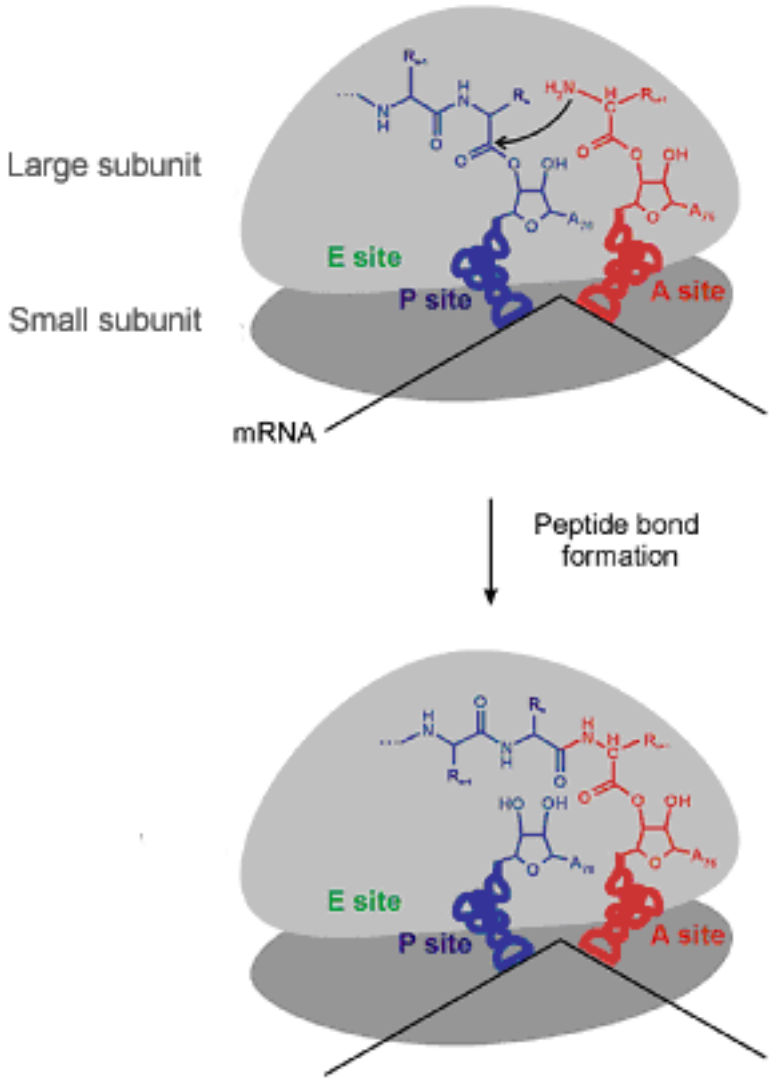


Figure 1.1 Schematic of peptide bone formation on ribosome, adapted from Beringer *et al*., (2007).

Extending from the PTC, there is a long aqueous tunnel through the large subunit named the ribosomal exit tunnel, where the growing nascent chain is temporally hosted and, after release from peptidy-tRNA, emerges to the cytoplasm (Yonath *et al*., 1987; Milligan *et al*., 1986). The exit tunnel is 100 Å long from PTC to the outside surface of the ribosome, and 15 Å in average diameter; in a roughly cylindrical shape, though

highly irregular in detail (Nissen *et al*., 2000) (Figure 1.2). The wall of the tunnel is mainly composed of rRNA components, but also includes regions of several ribosomal proteins which are unconserved between eukaryotic and bacterial ribosome. When first discovered in the mid-1980s, the ribosomal exit tunnel was considered as a rigid, passive tunnel. However, over the years, a considerable body of evidence has accumulated that is consistent with it being active during translation and that the exit tunnel participates in a range of co-translational events, such as crosstalk to the residues of the nascent chains or co-translational process of proteins (Ito *et al*., 2013; Ziv *et al*., 2005; Yap *et al*., 2009). The rproteins harboured inside the tunnel are indicted to be involved in these events (Ito *et al*., 2013). As 2A peptides only function with the eukaryotic translational system, the difference of rprotein components between the eukaryotic and bacterial ribosome might be considered one of the factors that drive this specificity, and candidates for studies that might lead to insight into its mechanism.

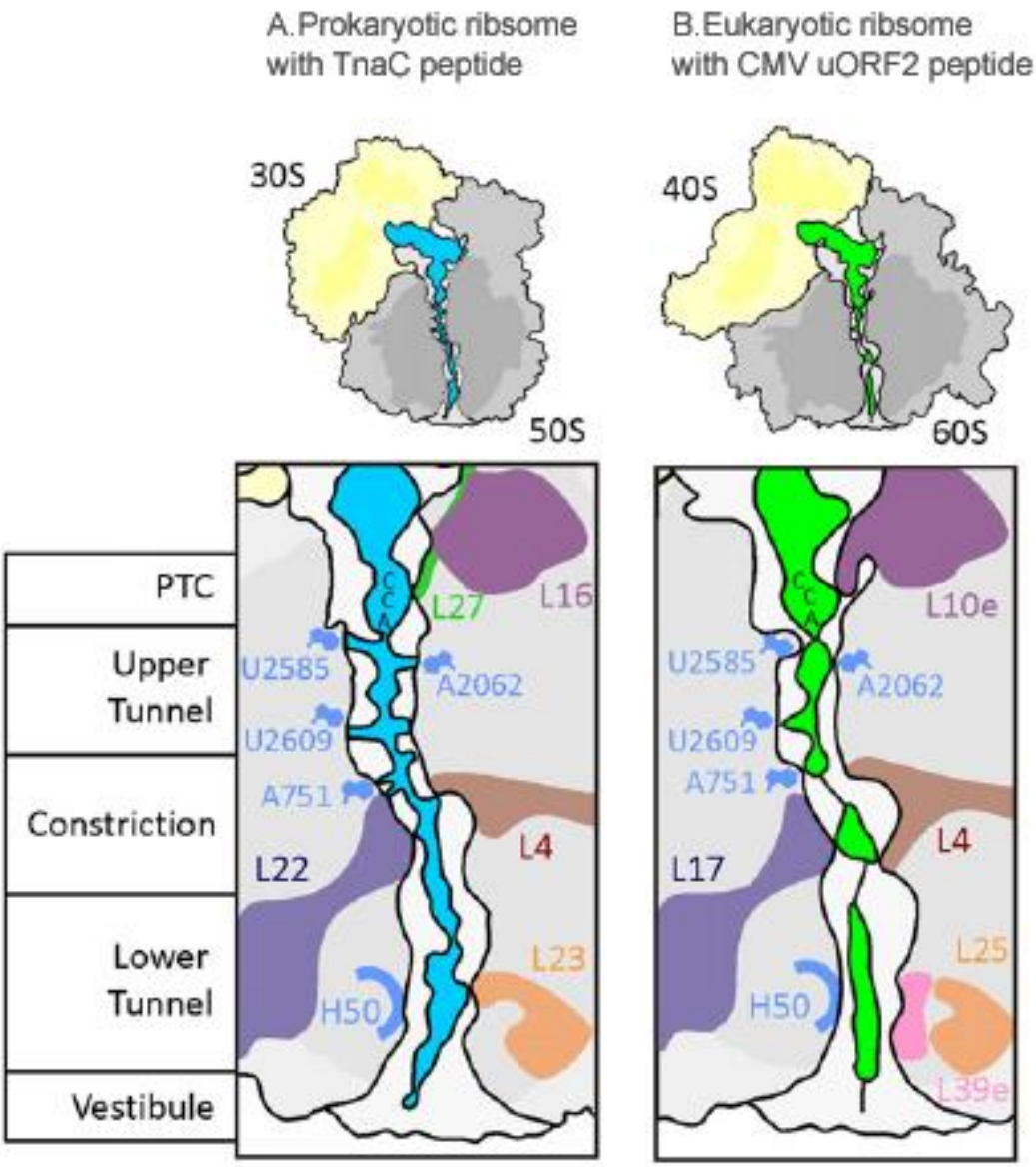


Figure 1.2 Schematic of the finding of cryo-EM of prokaryotic and eukaryotic ribosomal exit tunnels, with respective stalling peptides, TnaC (blue) and uORF2 (green) emerging from the ribosome. The residues of rRNA and rprotein are represented by azure and other colorful shapes respectively. Specifically, the azure curve H50 stands for the important rRNA Halices H50. Taken from Bhushan *et al*., (2010).

1.1.2 'Text-book' protein synthesis

Protein synthesis, the biogenesis of protein with mRNA as template in the ribosome, is a sequential and precise procedure, with several translation factors involved (for review see Schmeing *et al*., 2009 and Rodnina *et al*., 2009).

1.1.2.1 eukaryotic mRNA

The mRNA is the substrate of translation. In eukaryotic organisms, the basic structure of mRNA typically comprises 5 parts: a 5' 7-methyl-guanosine cap structure (m7G-cap), a 5' untranslated region (UTR), the coding region to be translated, followed by a 3' UTR and a poly(A) tail in the 3' end (Cogoni *et al*., 2000; Helm *et al*., 2006; Moore *et al*., 2005).

The 3' and 5' UTR may contain secondary structure like hairpin structure or internal ribosome entry site, which are often found to regulate translational efficiency or protein localization (Cenik *et al*., 2011). The m7 cap and poly (A) tail are necessary for the initiation of translation, as m7G cap offers the binding site of 43S pre-initiation complex, (see below) while poly(A) tail recruits poly (A) binding protein (PABP) that enhance the association between 43S complex and m7G cap structure. The coding region, or open reading frame (ORF), starts with the first methionine codon (AUG, termed as start codon) of the mRNA transcript, and ends with the one of three stop codons (UAA, UAG, UGA), with the ORF being decoded into a single protein chain (Acker *et al*., 2008; Rodnina *et al*., 2009).

1.1.2.2 Eukaryotic translational process

There are three major steps in translation: initiation, elongation and termination.

1.1.2.2.1 Initiation

The first and rate-limiting step in translation is initiation. Initiation entails two stages. During the first stage, eukaryotic initiation factors eIF3, eIF1A, eIF1 and eIF2 associated with initiator Met-tRNA and GTP, join the ribosomal 40S subunit, to form the pre-initiation 43S complex; on the other hand, eIF4E, eIF4A, eIF4G, the eIF4F complex, binds mRNA at its 5' m7 cap structure to facilitate the binding of the 43S complex to the mRNA. As the association of the 43S complex with mRNA relies on 5' m7 cap structure, this kind of mRNA recruitment is termed the 5'end dependent pathway. During the second stage of initiation, the 43S complex associates with the mRNA at its 5' terminus. It then moves on mRNA in 5' to 3' direction, 'scanning' for the initiation codon. On recognition of the initiation codon of mRNA, initiation factors release and ribosomal 60S subunit joins to the 40S subunit, forming the whole 80S ribosome; meanwhile the initiator Met-tRNA lands in the P-site of the ribosome, decoding the initiation codon and completing the initiation. There are three major differences between eukaryotic and prokaryotic initiation: firstly in prokaryotes the mRNA recruitment is 5'end independent, the association of pre-initiation complex and mRNA is mediated by the sequence of mRNA (termed Shine-Dalgarno) rather than mRNA binding factors; secondarily, the pre-initiation complex in prokaryotes binds directly, not slides to, the start codon, which is adjacent to the Shine-Dalgarno sequence of mRNA; thirdly, the fMet-tRNA, the initiator in prokaryotes, joins the pre-initiation complex after the complex binding to the mRNA (Amrani *et al*., 2008; Acker *et al*., 2008; Rodnina *et al*., 2009; Jackson *et al*., 2010).

1.1.2.2.2 Elongation

The second step of tranlation is elongation, the process of decoding the mRNA and extending the peptide chain. The basic cycle of elongation includes 3 processes.

In first process, a ternary complex comprising a eukaryotic elongation factor eEF1A, GTP and an aminoacyl-tRNA, enters the ribosome. This leads to a series conformational changes of the ribosome and induces the hydrolysis of the GTP by eEF1, and further results in the delivery of the cognate aminoacyl-tRNA to the A-site to decode the codon residing in the site. This codon recognition (decoding) is the key step of elongation, and one of the main translational progress that 2A changes. Therefore, in this introduction, a skeleton of the pathway of the decoding mechanism is offered. Due to the lack of detail on the decoding of eukaryotic system, this summary on decoding is based on studies carried out in bacteria. But because the elongation mechanism is considered highly conserved between prokaryotic and eukaryotic translation, the eukaryotic decoding presumably follows the same pathway as the prokaryote described below, albeit the decoding appears to take place somewhat more slowly and accurately in higher eukaryotes (Ito *et al*., 2013; Voorhees *et al*., 2013).

In bacteria, the ternary complex (in which the eEF1A replaced by EF-Tu, the bacterial homolog of eEF1A, that is also a GTPase) enters the ribosome, and the anticodon helix of the tRNA of the complex is delivered to the A site of ribosome, with assistance of rprotein L7/L12. In the A site, the tRNA anticodon and codon mRNA sequence pair to each others. During the pairing, three universal conserved residues of rRNA (in bacterial there are A1492, A1493, and G530 of 16S rRNA) interact with the minor groove of the cognate condon-anticodon helix at the first and second but not third positions, this monitors the correct base-pairing, The ‘cognate-specific’ binding energy generated from the these interactions is used to three events: (i) to increase the relative affinity of cognate tRNA (ii) to induce the large scale rearrangement of domains of small subunit

(closure conformation), which increases the fidelity of the translation; and (iii) to induce the conformational changes on the ternary complex, triggering the GTP hydrolysis by the EF-Tu, that further causes the release of the EF-Tu from the ribosome (Rodnina *et al*., 2005; Ogle *et al*., 2005). The hydrolysis of GTP also induces a proofreading step after the initial selection. As the conserved rRNA only monitors the first two pairs if the nucleotides of codon-anitcodon pairing, a fraction of tRNA with one mismatch to the codon at the A site remains in the initial selection. Some of the non-Watson-Crick base pairings are allowed in the decoding (wobble site of the genetic code at third position of the codon), but certain tRNAs with one mismatch (near cognate tRNA) are rejected in proofreading step, where whether the acceptor end of a tRNA is accommodated to the PTC is determined. The mechanism of proofreading remains largely unknown, but evidence suggests that it is a complex process involving the rRNA and rprotein of both ribosomal subunits and conformational rearrangement (Jenner *et al*., 2010; Demeshkina *et al*., 2012).

.

After the accommodation of the aminoacyl-tRNA comes the second process of elongation. The ribosomal peptidyl-transferase centre catalyses the formation of a peptide bond between the aminoacyl-tRNA in A site and peptidyl-tRNA (synthesized in previous elongation cycle) in the P-site, leaving the tRNA in P-site deacylated (see Figure 1.1) (Beringer *et al*., 2007).

The third and final process of elongation is translocation: after peptide bond formation, another elongation factor eEF2, a GTPase, enters the ribosome with a GTP, and catalyzes the GTP hydrolysis there. With the energy released from the hydrolysis, the ribosome translocates 3 nucleotides toward 3’ end of mRNA, moving the peptidyl-tRNA in A site and deacylated tRNA in P site to the P site and E site respectively. Therefore, the A site is emptied, ready for the next elongation cycle (Taylor *et al*., 2007).

1.1.2.2.3 Termination

The third and last step of translation is termination, in which the ribosome arrives the one of three stop codons (UGA, UAG, UAA) that ends the open reading frame in the mRNA and the nascent polypeptide is released. In eukaryotic translation, two eukaryotic release factors are required for termination: eRF1 and eRF3. These two factors enter the ribosome in a state of complex (Pisareva *et al*., 2006).

Unlike the prokaryotic release factors that use different factor to decode different stop condon, the eukaryote recognizes all three stop codons with eRF1 only. Studies have revealed that the conserved motifs of the N-terminal portion (domain 1) of the eRF1, such as NIKS-loop or YxCxxF motif, play the centre role of stop codon recognition. The concomitant amino acid replacement(s) on these motifs may cause the impairment on the efficiency of the recognition, or produce the 'UGA only' eRF1 mutants that can only decode stop condon UGA. These findings suggest that the eRF1 N-domain recognize stop codon through complicated three dimensional network formed by its conserved resides (Chavatte *et al*., 2002; Frolova *et al*., 2002; Seit-Nebi *et al*., 2002; Kolosov *et al*., 2005; Cheng *et al*., 2009).

The eRF1 is very versatile during termination. Except its N-terminal portion that recognizes all three stop condons, it also contacts several key nucleotides of peptidy-transferase centre with the conserved GGQ motif in its domain 2. This realizes the second major function of the factor, triggering the release of nascent chain from peptidyl–tRNA.

The eRF3, a GTPase, binds to eRF1 in high affinity with the domain 3 of the latter (Kolosov *et al*., 2005; Cheng *et al*., 2009). Studies suggest, the hydrolysis of GTP catalyzed by eRF3, induces the conformational rearrangement of eRF1, resulting the couple of two major functions of eRF1: stop codons recognition and catalysis of peptide relaease (Alkalaeva *et al*., 2006).

Although the eRF1 and eRF3 plays the central role of the translational termination, recent studies have revealed the involvement of other factors. For example, the iron-sulphur protein RNase L inhibitor (Rli1) was reported associated physically with eRF1 during termination, and regulation of the Rli1 affected the efficiency and accuracy of the latter with a positive related manner (Khoshnevis *et al*., 2010). Besides, the Rli1 was also found collaborating with released factors to slip the ribosomes into their subunits after the release of the nascent chain, thus trigging the recycle of the translational components (Pisareva *et al*., 2010).

1.2 Unconventional translational mechanism

Section 1.1 describes the course of standard 'textbook' translation. But the progress of protein synthesis is not always as simple as this description. Occasionally the outcomes of translation do not match those predicted from inspection of mRNA sequences, and these events are termed 'translational recoding events'. In other cases, the efficiency and output of translation, can be co-translationally controlled by regulatory factors, with the approach of pausing or stalling the course of translation. As mentioned above, these unconventional events are driven by either the template mRNA or product nascent peptide chain.

As 2A reaction, the topic of this project, is a translational recoding event, and directed by the peptide sequence, this section reviews several unconventional events, to look for the insights for understanding the 2A mechanism.

1.2.1 *Internal Ribosome Entry Sites (IRES)*

As mentioned in section 1.1.2, in the typical eukaryotic initiation, recruitment of the mRNA is via 5'-end cap dependent pathway. In some cases, however, the structure of

RNA is able to recruit the ribosomal subunits directly and initiate translation in a 5'-end independent manner. Such RNA structures are collectively termed internal ribosome entry sites (IRESs). IRESs were firstly found in virus study, but later also in eukaryotic cellular mRNA. As of 2009, there are 68 viruses and 115 eukaryotic mRNA reported to contain IRES (www.iresite.org). IRESs in different organism varies in biological role. For eukaryotes, IRES dependent initiations are active during cellular stress, mitosis or apoptosis conditions, where the canonical factors that typical initiation needs compromised or impeded. This suggests cellular IRESs are developed to deal with the transient stress (Holcik *et al*., 2000; Stoneley *et al*., 2000; Lang *et al*., 2002). The IRESs of virus, on other hand, are suggested playing a role of regulatory element. Under the infection condition, the canonical initiation is prevented by the virus with proteolysis of initiation factors, or host cell, with regulation. In such condition, the virus needs IRESs-induced initiation to start the translation of viral proteins. Studies of IRES suggest that there is no universal mechanism of IRES. In terms of origins, mechanism and sequence similarity, the reported viral IRESs can be classified into 4 types. Each type requires unique subset of initiation factors and cellular proteins that have no known relatives with translation (Redondo *et al*., 2012; Yu *et al*., 2000). However, one unifying feature is that all these mechanisms include the unconventional interaction between the mRNA structure of IRES and normal translational apparatus (Jackson *et al*., 2013).

1.2.2 Ribosomal frameshifting

One of the most important functions of the ribosome is to maintain the reading frame. However, in specific situations ribosomes are prompted to move into a different reading frame (both forward, marked +, or backward marked -), to produce different proteins from the same transcript.

Ribosomal frameshifting occurs in many situations. For example, many retroviruses use this mechanism to produce Pol protein (frameshifting -1) as a fusion to their Gag protein (Brierley *et al*., 2006; Dos Ramos *et al*., 2003). In bacteria and eukaryotes, the

expression of type II release factor (RF2), and ornithine decarboxylase antiyme are mediated by +1 frameshifting respectively (Craigen *et al*., 1987; Ivanov *et al*., 2004).

The typical elements of mRNA that induce-frameshifting are two sequences: the first is a so-called slippery sequence that constitutes the shifting site. In such a sequence, frame-shifting does not affect the condon-anticondon pairs of tRNA with mRNA in the P and A site of the ribosome. For example, the shifting site of Gag-Pol gene of HIV-1 is UUUUUUA, before the shifting the sequence is read as UUU UUU A, and after shifting, U UUU UUA. In both frames Phe-tRNA is able to pair with the codon in the P and A sites cognate to, or form a wobble site (see section 1.1.2.2.2) with the mRNA sequence (Mouzakis *et al*., 2013). The second element is the secondary structure of the sequence downstream to the shifting site, usually a pseudoknot, a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another (Giedroc *et al*., 2009; Brierley *et al*., 1989). It is general believed, that the frameshifting is achieved by the *cis* coupling of the slippery sequence and structured mRNA. The current model to explain the frameshifting mechanism (Dinman *et al*., 2012) comprises 2 major steps: firstly, while the slippery sequence arrive to the P and A site of ribosome, the ribosome was paused by the resistance of the downstream structure; and secondly, a fraction of ribosomes slip 1 nucleotide to 5’ or 3’-direction, but the exact point of this slip has not been identified yet (Giedroc *et al*., 2009; Brakier-Gingras *et al*., 2012; Plant *et al*., 2003; Liao *et al*., 2012). A recent cryo-EM study generated data consistent with a mechanical explanation for how a pseudoknot in the RNA can cause a -1 frameshift. In this study, the slippery sequence was replaced with a non-slippery sequence. The structure of ribosomes translating an mRNA containing this modified RNA suggested that the pseudoknot became wedged into the ribosomes, causing them to stall, unable to complete the translocation step in the elongation cycle. It was inferred from this that in the presence of the slippery sequence, the tension within this complex could be released by the -1 frameshift on the mRNA (Namy *et al*., 2006).

1.2.3 Stop codon readthrough

Translation termination occurs when one of three stop condon occupies the A site of ribosome. However, a growing number of events have been identified where tRNAs (termed suppressor tRNA) are able to decode the stop codon (nonsense condon), allowing translation to proceed downstream. Such events are together termed 'programmed stop codon readthrough'. As during the readthrough event, there still are a fraction of ribosomes terminating the translation at the stop codon, the event lead to production of 2 proteins: the product that released at the stop codon (termination product), and the readthrough product, from the single transcript. Many viruses employ this mechanism, and an example is the RNA phage Qβ infecting *E. coli*. During the translation of the end of Q-coat protein, a fraction of ribosomes incorporate tryptophan instead of terminating translation at a UAG stop codon, resulting a elongated coat protein that is critical for its infection (Hofstetter *et al*., 1974); a second example is the murine leukemia virus, (MuLV) that generates its Gag-Pol precursor, by decoding its UAG termination condon with $tRNA^{Glu}$ (Yoshinaka *et al*., 1985; Yoshinaka *et al*., 1985(2); Brierley *et al*., 2007). Though the majority of readthroughs identified are in viral genes, there are several cellular genes employing the mechanism. For example, *Drosophia* uses readthrough to regulate the ratio of two proteins, Hdc and the long Hdc protein, during development (Steneberg *et al*., 2001).

The 3'- adjacent nucleotides to the stop codon are believed to be important for readthrough. Taking into account to the requirements of the mRNA sequence and/or structure stimulating the event, readthrough occurring in eukaryotic system can be

classified into three types: type I, with a stimulatory motif ‘UAG-CAA-UYA’; type II, the three nucleotides 3’-ajacent to the stop codon UGA are either CGG or CUA; and type III (represented by MuLV mentioned above), in which the high-level of readthrough requires the secondary structure of mRNA several nucleotides downstream from the stop codon (Brierley *et al*., 2007; Beier *et al*., 2001; Harrell *et al*., 2002). However, some recent studies point out, in the first two type of readthrough, downstream structured mRNA is more generally involved in the events than previously thought, and several cases of readthrough are reported relying both 3’- adjacent nucleotides and the secondary structure of downstream mRNA (Napthine *et al*., 2012; Firth *et al*., 2011).

1.2.4 Ribosomal stalling peptides

As mentioned in section 1.1.1, the ribosomal exit tunnel plays an active role in the biogenesis of proteins. The interaction between nascent peptide chain and components of exit tunnel is perhaps is a frequent occurrence. In some cases, this kind of contact induces conformational changes to the ribosome and/or the positioning of the nascent chain, and results in stalling of translation, with specific codons in the PTC. Peptide-induced translational stalling in different origins is evolved independently and carries out distinct affects (Ito *et al*., 2013; Yan *et al*., 2010). They are not translational recoding, but since the 2A reaction includes a translational pause (see section 1.3), it is an important aspect to translation to discuss here. This section offers an introduction to four typical ribosomal arresting peptides (RAP), and a more comprehensive list of known RAPs is provided in Table 1.1.

RAP	Organism	Arrest sequence (P site underlined)[a]	Inhibition point
SecM	*Escherchia coli*	FSTPVWISQ AQGIRAGP	Peptidyl transfer
SecM	*Mannheimia succiniciproducens*	HAPIRGSP	
MifM	*Bacillus subtilis*	RITTWIRKVFR MNSPVN*DEE DAGS*	
ErmCL	*Staphylococcus aureus* (*pE194*)	IFVI	
ErmAL1	*S. aureus* (*Tn554*)	IAVVE	
Cat leader	*B. subtilis* (plasmid)	VKTD	
CmlA leader	*E. coli*	KNAD	
TnaC	*E. coli*	WFNIDNK IVDHRP*	Termination
TnaC	*Proteus vulgaris*	WFNIDSEL AFFFP	Peptidyl transfer
AAP	*Neurospora crassa*	TSQDYLSDHL WRALNA*	Termination
CGS1	*Arabidopsis*	RRNCSNIGVA QIVAAKWS	Translocation
AdoMet DC uORF	Eukaryotes	DIS*	Termination
OAZ1	*Saccharomyces cerevisiae* (Eukaryotes)	Large domain	
***UL4* uORF2**	Cytomegalovirus	SAKKLSSLLT CKYIPP*	Termination
2A peptide	Foot-and-mouth disease virus	LKLAGDVESNP G(P)	Peptidyl transfer
XBP1u	Human	YQPPFLCQW GRHQPSWKPL MN	
Poly Lys/Arg	Eukaryotes	K_{12}, etc.	
WPPP	*E. coli* (Lab evolved)	FQKYGIWPPP	Peptidyl transfer

Table 1.1 Main translational arresting peptides, their organism, arresting sequence (motif highlighted with red colour) and inhibition point. Taken from Ito *et al*., (2013).

1.2.4.1 secM

Secretion monitor (*secM*) lies upstream to, and regulates the expression of *secA* gene, encoding SecA, a key component of protein secretion apparatus of *E. coli*. I SecM has both an N-terminal signal sequence, and a sequence (FxxxxWIxxxxGIRAGP) that causes ribosomes to stall. This in turn disrupts the secondary structure of the *secM-secA* mRNA, which, interestingly, leads the entry of new ribosome that translates *secA*, and finally results the upgrade of SecA expression (Nakatogawa *et al*., 2001). Efficient engagement of nascent SecM with the export machinery counteracts the effect of the stalling peptide, and translation efficiency of SecA is lowered. This provides a regulation of SecA (which encodes an ATPase that drives export through the secYEG translocon) synthesis in response to translocation efficiency.

The short peptide SecM (F^{150}xxxxWIxxxxGI**RAGP**^{166}) directing the ribosomal arrest is from amino acids 150 to 166 of the 170-amino-acid polypeptide. During the stall, the codons of Pro166 and Gly165 are positioned in the A and P sites respectively, and Prolyl-tRNA also lands in the A site. However, the ribosomal peptidyl-transferase fails to form the peptide bond between Prolyl-tRNA and peptidyl-tRNA in P site (Nakatogawa *et al*., 2002; Muto *et al*., 2006; Garza-Sanchez *et al*., 2006).

Insertion of fluorescent probes into the SecM peptides allowed FRET (fluorescence resonance energy transfer) studies to be carried out on the stalled SecM-ribosome complex, and the data collected reveal that during the reaction, all the 17 amino acids of the motif rest in the exit tunnel, in a compact, perhaps helical form, and mutagenesis on the ribosomal components that consistent the narrowest region of the exit tunnel, both

of rprotein (L22) and rRNA, (23S rRNA) compromise the SecM-induced stalling significantly. These facts strongly suggest the existence of interactions between the SecM nascent chain, and the exit tunnel component (Nakatogawa *et al*., 2002; Woolhead *et al*., 2006). Cryo-electron microscope studies reveal the distinct rearrangement to the ribosome structure and peptidyl-transferase centre, which is potentially the result of such interactions (Mitra *et al*., 2006; Bhushan *et al*., 2011).

The SecM motif (F^{150}xxxxWIxxxxGI**R**A**GP**^{166}) composed of nine amino acids was isolated by alanine-scanning mutagenesis (Nakatogawa *et al*., 2002). However, a further mutagenesis study revealed that, besides the G-P pair at the C-terminal end, only arginine at position 163 is indispensable for the full function of SecM, and the mutations at other key amino acids could be compensated by re-modeling of other positions. This finding suggests while the Arg163 appears to play singular function in translational stalling, the other residues between 150 and 165 that play a "secondary" role probably fall into at least one other functional class (Yap *et al*., 2009). Considering the research discussed above that suggests the interaction between nascent chain and ribosomal component, a possible explanation is that, all residues located at 150~166 except Arg163, by forming a certain conformation or interaction with the exit tunnel, correctly position Arg163 into a precise intra tunnel site where it is able to contact critical component hence further trigger the ribosome stall, with the G-P pair of the peptide at the decoding site. This explanation was a consistent with the cryo-EM study on the peptides, which offered critical evidence that R163 contacted with A2603 of ribosome (Bhushan *et al*., 2011).

Proline at the A site itself is poorest electrophile for peptide bond formation (Nathans *et al*., 1963) and reaction involving it are then perhaps more likely to be sensitive to the perturbation of ribosomal geometry than other amino acid. The spatial stack of its unique proline ring structure also means that it is less flexible to accommodate the subtle ribosomal rearrangement. The glycine at the P site is also poor electrophile (Nathans *et al*., 1963), making G-P one of the worst couples in term of making the peptide bond (Rychlík *et al*., 1970). This situation, plus the ribosomal conformational

change potentially caused by the nascent chain-ribosomal exit tunnel contact, might be the reason that peptidyl-transferase fails to form the peptide bond between the final Prolyl-tRNA and peptidyl-tRNA in P site (Yap *et al*., 2009).

1.2.4.2 TnaC

The leader peptide of the tryptophanase operon of *E. coli*, TnaC, is a 24 amino acids long peptide. It responds to high concentrations of tryptophan, upregulating the expression of downstream genes whose function is catabolism of tryptophan, with a delicate mechanism starting with a ribosome arrest (Gong *et al*., 2001; Gong *et al*., 2002).

At the point that the ribosome is stalled by the peptide, the P and A sites are occupied by the final proline codon (Pro24) of the peptide open reading frame, and the stop codon in the end of the ORF respectively. However, termination at the stop codon in A site does not occur, thus, the ribosome is arrested by TnaC peptide at its own stop codon. Mutagenesis studies have highlighted key residues of TnaC, they are Trp12, Asp19 and final proline in the 24th position (Cruz-Vera *et al*., 2008). Similar to SecM, mutations in components of the ribosomal exit tunnel constriction suppress TnaC-induced stalling. Like SecM, this again agrees with the contact of nascent chain and exit tunnel (Cruz-Vera *et al*., 2005). The possible result of these contacts is revealed by a recent cryo-EM research, that indicates in the arrested ribosome, two conserved nucleotides of PTC adopt conformation that precludes release factor binding (Seidelt *et al*., 2009). These findings/data explain the failure of translation termination at the TnaC stop codon.

1.2.4.3 Cytomegalovirus UL4 uORF2

The ribosome arrested peptides reviewed in last two sections are functional in bacterial translational apparatus. Similar phenomena have been observed in eukaryotes. UL4 uORF2 of cytomegalovirus, a virus infecting human, is a case in point (Cao *et al*., 1996). The UL4 coding region is present on a transcript containing three upstream open reading frames (uORF). The second one (uORF2), a 22 amino acid long leader peptide, suppresses translation of the downstream main open reading frame with a mechanism that, like TnaC, is due to pausing of ribosomes with the stop codon of the uORF in the ribosomal A site. This results in a persistent covalent attachment between the nascent chain and the final tRNA, which, also similar to TnaC, is proline (Cao *et al*., 1996; Cao *et al*., 1995).

Unlike TnaC that precludes the release factors from ribosome by altering ribosomal conformation, during the uORF2-induced ribosome stalling, eRF1 (but not eRF3) is recruited to the ribosome, and mutations in GGQ motif of eRF1 significantly impair the duration of the stall. This finding suggests the release factor eRF1, particularly the domain 2 that catalyze the hydrolysis of nascent chain in canonical translation, play a critical role in the arrest. Furthermore, the final two prolines of uORF2 are required for the stalling (Janzen *et al*., 2002). To explain these data, a model was proposed that, the carboxy terminal proline residues, and the GGQ motif of eRF1 contacted each other to stabilize an intermediate in termination, while the potential interaction(s) of ribosomal component and the rest of nascent peptide altered the geometry of ribosomal PTC and/or shift the position of uORF2-tRNA, to enable such contacts (Figure 1.3) (Janzen *et al*., 2002).

This model was partly supported by a recent cryo-EM study, which showed direct interactions of the nascent chain and the components of exit tunnel (rprotein L4 and L17), and significant changes at the PTC: the eukaryotic-specific rprotein L10e contacted uORF2-tRNA at its CCA end (Bhushan *et al*., 2010). But more details of this type of termination arrest will undoubtedly emerge from further analysis.

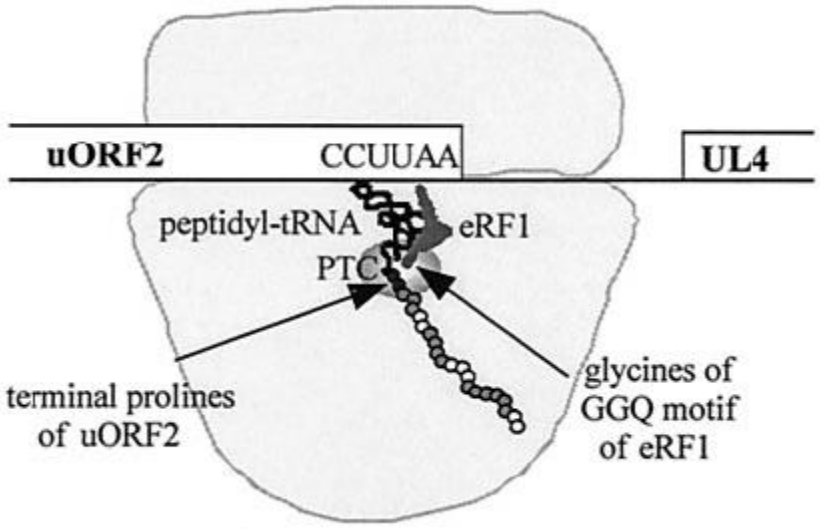


Figure 1.3 Model of uORF2 induced termination inhibition, taken from Janzen *et al*., (2002). The altered PTC allows the GGQ motif of eRF1 contacting with final prolines of the peptide, stabilize an intermediate in termination.

1.2.4.4 Arginine attenuator peptide (AAP)

Another typical ribosomal arresting peptide working with eukaryotes is arginine attenuator peptide. During their translation, a variety of fungal species including *S. cerevisiae* and *N. crassa* use AAP uORF arresting the ribosome, to restrict translation of a component of arginine biosynthetic pathway downstream, in response to high arginine level (Fang *et al*., 2000).

Like UL4 uORF and TnaC described above, AAP arrests the ribosome with its stop codon in the A site, but the final amino acid condon that is in P site during the stall, is not a conserved position of the peptide. Therefore, the motif of AAP is in the 'middle' of the peptide, with non-conserved amino acids forming its N- and C-terminal portion. Most intriguingly, the stop codon in the end has been proved unnecessary for its ability to stall ribosome: replace it with a sense condon, ribosome stall at several positions just downstream to the AAP. The mechanism of AAP-induced ribosome stalling remains unknown, but according its unique features, it would appear to be very distinct from other three peptides introduced in this section (Fang *et al*., 2004; Wang *et al*., 1997).

1.3 2A sequences and the 2A reaction

A focus of research in the Brown laboratory for the past few years has been short peptides, collectively known as 2A peptides, and also termed 'CHYSELs' (cis-acting hydrolase elements) (de Felipe *et al*., 2004) that both pause ribosomes and dictate a translational recoding mechanism: They cause a co-translational 'skip' in peptide bond formation at their C-terminus (Doronina *et al*., 2008). This results in separation of the polypeptide containing the 2A sequence, into two parts: the product containing the sequence upstream to and including most of 2A, (upstream product) and product downstream to and including the final proline of the peptide (Downstream product). The activity of 2A sequences is only dependent on the peptide sequence, not RNA sequence (Ryan *et al*., 1994).

2A peptides have been found in a number of viruses, where they play important roles in their life-cycle and reproduction, as well as cellular genes of sea urchin (M. Ryan personal communication) and in Trypanosomal repeat sequence (Heras *et al*., 2006; Oden *et al*., 2013). Typically, the 2A peptide has length of ~20 amino acids. The sequences can be divided into two parts: an N-terminal non-conserved sequence which in some 2As has a strong α-helical propensity, followed by a conserved motif D(V/I)EXNPGP. The separation of the nascent chain is at the C-terminal end of the conserved motif, between glycine and proline (Figure 1.4).

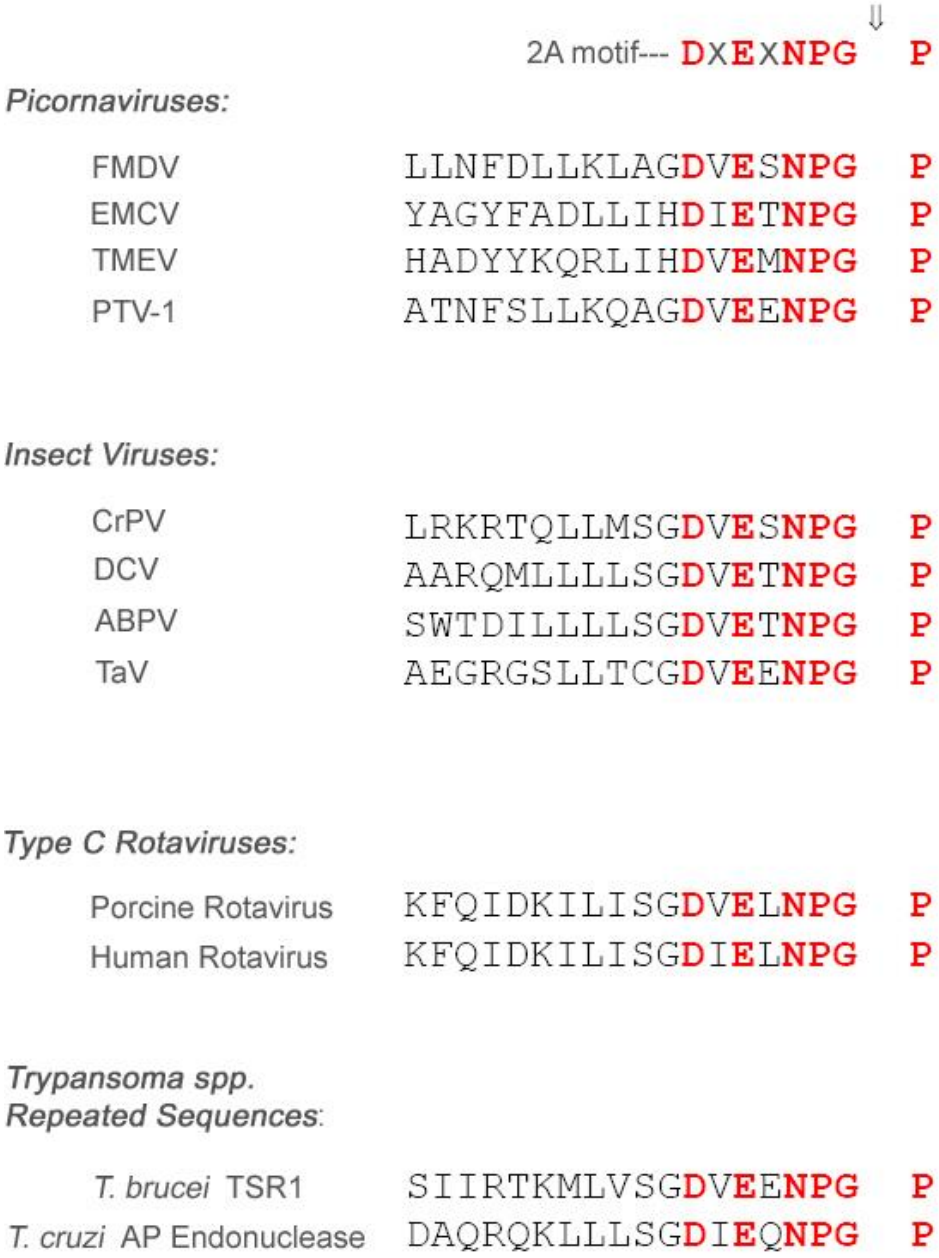


Figure 1.4 2A peptides: The alignment of 2A-like peptides sequence from different species shows a conserved motif D(V/I)EXNPGP at C-terminus. Data collected from http://www.st-andrews.ac.uk/ryanlab/Index.htm

1.3.1 The 2A reaction is a co-translational event rather than a proteolytic cleavage

The first 2A peptide identified and the most extensively studied so far is the 2A peptide of the *aphthovirus* foot-and-mouth disease virus (FMDV). The studies on 2A peptide can be traced back to the work of Ryan and his colleagues in 1991 (Ryan *et al*., 1991). In that research, it was found that a 19 amino acide long short peptide spanning the 2A region of the foot-and-mouth disease virus polyprotein, was responsible for a 'primary proteolytic processing' event occuring between the 2A and 2B regions of the polyprotein. Initially it was thought that the separation of the polypeptide chain was carried out post-translationally and that 'either an unidentified virus-encoded proteinase may be responsible, or that 2A acts as a substrate for a host cell proteinase', as 2A protein of many other viruses is a protease. However, in follow-up work (Ryan *et al*., 1994) this hypothesis was discounted.

Key observations in (Ryan *et al*., 1994) were made with an artificial polyprotein system (plasmid pCAT-2A-GUS), in which the 19 amino acid 2A peptide from FDMV was inserted between two reporter protein, chloramphenicol acetyl-transferase (CAT) and β-glucuronidase (GUS). In the *in vitro* translation experiment with the rabbit reticulocyte lysate system, high-level of CAT-2A (upstream product) and GUS (downstream product) was observed, as well as a small amount of full-length product (CAT-2A-GUS). This means existence of the high-level 2A-mediated 'cleavage' activity. Given the 2A peptide is the only component in the experiment from FMDV, this result discounted the possibility that virus-encoded proteinase participated in the 2A reaction. Another important result within this work, kinetic analysis, showed that the full-length product was stable, arguing against the involvement of host cell proteinase in a post-translational cleavage of the protein. Therefore, the suggestion that a co-translation event resulted in 'cleavage' of the protein was suggested for first time (Ryan *et al*., 1994).

Further evidence in favour of the co-translational model of 2A reaction was a molar imbalance of the up- and down-stream products, with a greater amount of upstream product being made than downstream seen in some experiments. This was observed with several independent fusions containing the FMDV 2A peptide. As this imbalance could not be explained by protein degradation or premature termination of transcription or translation, the only possibility was that it was due to the differential rates of synthesis of the two products, for example, the ribosome might fall off from mRNA after generation of upstream production. Therefore, the 2A-mediated cleavage is strongly suggested as a novel translational event within ribosome, rather than proteolytic reaction after translation (Donnelly *et al*., 2001a; Donnelly *et al*., 2001b).

More recent and decisive evidence confirming the 2A reaction as a co-translational event has come from study in Dr. Brown's laboratory (Doronina *et al*., 2008). In this study, it was found that, with truncated transcripts only extending to the Pro19 of 2A (without stop codon), the 2A reaction still occurred, the peptide driving the ribosome to release the upstream product. This result demonstrated conclusively the co-translational nature of the 2A reaction, and that the reaction takes place in PTC of the ribosome.

1.3.2 Features of 2A reaction

During investigation into the primary mechanism of the 2A reaction, many other faces of it were discovered. In the study of Donnelly *et al*., 1997, it was found that the upstream context of 2A in FMDV, although not crucial for 2A reaction, played a important role, as insertion of upstream FMDV capsid protein 1 D sequences (which preceed 2A in the virus) upstream of 2A increased activity. The same study also discovered that, while the 2A functions in a variety of eukaryotic translation systems, it was not active when translated by prokaryotic ribosomes.

1.3.3 Beyond FMDV 2A

While the co-translational mechanism of the FMDV 2A reaction was being uncovered, many other examples of '2A-like' short peptides had been found and the common features of these peptides had been noted. Donnelly *et al*., 1997 found that the homologues of 2A in other members of family *picornaviridae,* namely Theiler's murine encephalomyelitis virus (TME) 2A protein and encephalomyocarditis virus (EMC) 2A (both cardioviruses), had the same autonomous cleavage activity as FMDV 2A. Analysis of these sequences revealed a common motif of this kind of elements: a non-conserved N-terminus with strong α-helical propensity (which may be important for 2A reaction) followed by a conserved motif D(V/I)EXNPGP at the C-terminus of the peptide. As more virus genome sequences have become available, the number of 2A-like peptides increased, as has the range of viruses known to have these elements (Reviewed in Luke *et al*., (2008) and Sharma *et al*., (2012)). Further, studies also revealed that some cellular genomes contain '2A-like' peptide, such as the ones of *Trypansoma brucei* and *Strongylocentrous purpuratus* (Donnelly *et al*., 2001b and Heras *et al*., 2006). These peptides were tested in artificial polyprotein systems and confirmed in many cases to be active.

Beside research into requirements for and mechanism of the 2A reaction, the possibility of putting FMDV, and other 2A-like peptides to work as a tool had also been considered (de Felipe *et al*., 1999; de Felipe et al.2000; de Felipe *et al*., 2002; de Felipe *et al*., 2003; de Felipe *et al*., 2004; Liu *et al*., 2007; Szymczak *et al*., 2012). For example, in a study the 2A peptides (from FMDV and other viruses) was used to build polycistronic mRNAs that can be used in gene therapy delivery vectors (de Felipe *et al*., 1999).

1.3.4 The mechanism of 2A reaction

Although 2A peptides are increasingly used as research tools and are even being trialled in clinical applications (de Felipe *et al*., 1999; Szymczak *et al*., 2012) the mechanism of the 2A reaction is largely unclear. Current data and a model for the reaction based on these are outlined below.

1.3.4.1 A ribosomal stall in the C-terminus of 2A

Similar to the stalling peptides such as that encoded by the uORF2 and the secM peptides, 2A and 2A-like sequences pause the ribosomes. The existence of pausing at 2A was firstly suggested by an *in vitro* study in which translation of a 2A-containing protein was carried out in presence of low concentration of puromycin (Donnelly *et al*., 1997). Puromycin is an analogue of aminoacyl-tRNA which is randomly added to the C-terminus of nascent proteins during the translation elongation, independent of the nature of the codon present in the ribosomal A site. Inclusion of puromycin throughout translation should result in its incorporation throughout the synthesis of the polyprotein and lead the (random) puromycin-terminated truncation of translation products. But in the study a significant higher level of puromycin-terminated products (identified by immunoprecipitation with anti-puromycin antibodies) of the size of upstream products

of the 2A reaction was detected, which was suggested to reflect ribosomes staying at the C-terminus of 2A for a longer time than the other positions.

Further evidence of a pause in translation was collected in a more recent study (Doronina *et al*., 2008). using the primer extension inhibition (toeprint) technique, in which reverse transcriptase is used to map the positions of ribosomes and other factors on RNA. When *N. crassa* translation extract (a very sensitive system for toeprint analysis) was programmed with mRNA encoding a 2A sequence, a toeprint signal at +16/17 nucleotides to the first base of glycine 18 condon of 2A was observed. Given that the toeprint of eukaryotic ribosomes map 15 to 17 nucleotides distal to the first base of the codon in the P site due to their size, this result not only supports the finding of existence of a pause but also indicates the position of the ribosome when the pause occurs - namely when the glycine 18 and proline 19 condons are in the P site and A site, respectively.

On the other hand, however, the same study also revealed that the strength of the toeprint signal at 2A is not as strong as at other nascent peptides that induce translational stalling, like the *AAP* leader peptide which was used as a control in these experiments. These data, combined with results from time course *in vitro* study of 2A where the downstream products appeared at the similar time as full-length products, suggests that this pause may not be as long as the samples of other short peptide-mediated translation pause where ribosome can be blocked for minutes. (Doronina *et al*., 2008b).

1.3.4.2 The involvement of eukaryotic release factors in 2A reaction

As described in section 1.1.2, protein translation is a sequential and precise procedure with several translation factors involved. Therefore, the question of whether the 2A reaction, as a co-translational event, is mediated by the translation factors or affects the normal action of these factors, is important to discerne.

For uncovering factors that play a role in the 2A reaction (Doronina *et al*., 2008), expressed a reporter containing 2A in a range of *S. cerevisae* strains in each of which the activity of a specific translation factor is altered. It was found that over-expression of the 2A-containing protein in strains in which eRF1 or eRF3 factors were compromised, led a growth-inhibition, and this was observed neither in wild type cells nor RF compromised cells with no or low 2A expression. Further experiments revealed that the readthrough rate of a separate reporter of stop codon read-through co-expressed in the same cells increased with the expression of the wild 2A but not inactive 2A mutant. This suggests that availability of RF or its activity is reduced by 2A expression. These results suggest that ribosomes at 2A may recruit RF while the ribosome pause at 2A and leave insufficient RF for normal translation. However, the fact that the 2A induced translational pause is comparatively short weakens this argument (Doronina *et al*., 2008).

The effects of compromised RF on the 2A reaction were also investigated in the same study. The results gained in *in vitro* and *in vivo* experiments revealed that, while reduced eRF1 activity led the 2A activity declining, compromised GTPase activity of eRF3, interestingly, increased the proportion of 2A upstream products and decreased both of those of full-length and downstream products. Considering the conventional role of eRF3 in translation, one possible explanation of this effect is that in 2A reaction, eRF3, a GTPase, assists eRF1 released from the ribosome and therefore compromised eRF3 may induce a longer time during eRF1 combining with ribosome and yielding more upstream products (Doronina *et al*., 2008).

1.3.4.3 The outline of the model of 2A reaction

Summarizing the data obtained so far, a model of the 2A reaction, comprising 3 steps can be proposed (Figure 1.5).

The first step is the translation of upstream product and the 2A peptide up to and

including the penultimate amino acid, glycine. The ribosome then pauses, presumably (like other peptides arrested ribosome) due to the interactions between the ribosome exit tunnel and the 2A peptide, leaving the penultimate glycine and final proline codon in the P and A site of the ribosome respectively. Due to some as yet undefined reason, (perhaps the specific spatial conformation of the peptide and tRNAs), the peptide bond between glycine and proline is not formed.

The second step is the hydrolytic release of upstream product from tRNA. As mentioned above, there is clear evidence (Doronina *et al*., 2008) that release factors are involved in this process, but how 2A recruits these factors and their specific role in the reaction is still not entirely defined.

The third step of 2A reaction is restart of translation from C-end proline of 2As, this territory still remains uncharted, and no factors have been identified that influence this step in the reaction.

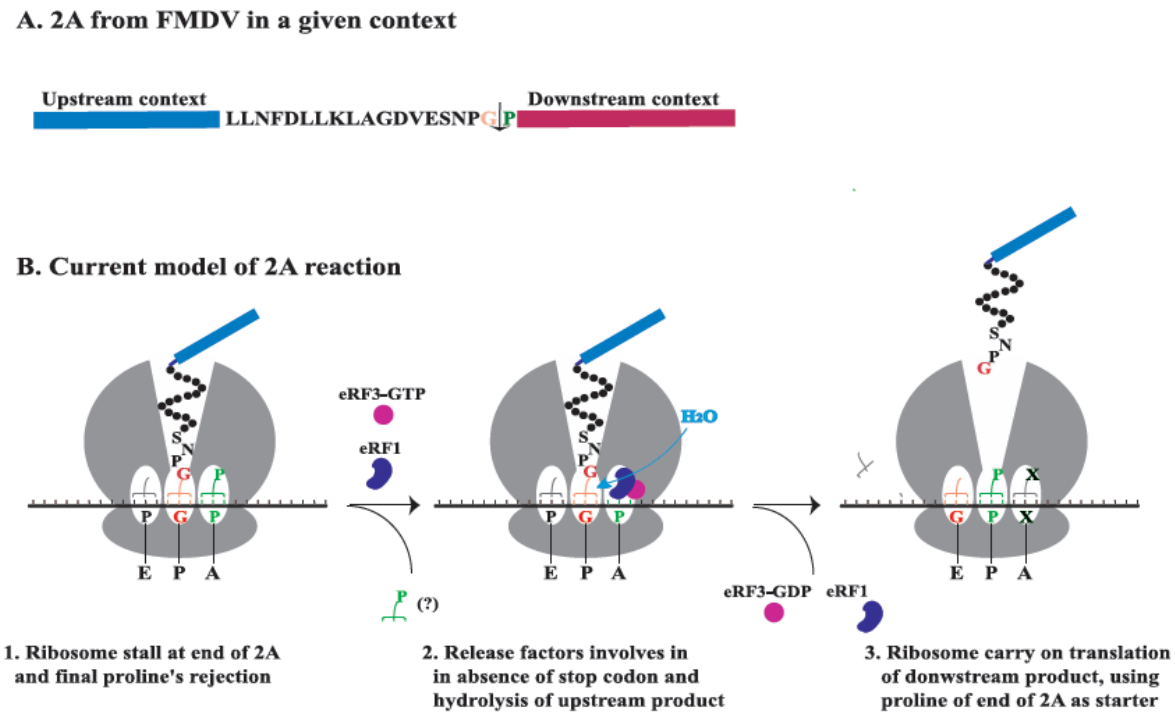


Figure 1.5 Outline model of 2A reaction. See text for details.

1.4 Aims of present study

The general aim of the study presented in this thesis was to extend the knowledge of the mechanism of 2A reaction. This was realized by two aspects of research:

1) A site-specific mutagenesis study on the 2A sequence: the aim of the site-specific mutagenesis was to analysis the functional element of 2A sequence. In the studies before my project, a random mutagenesis study was employed to analyze the 2A sequence, here my aim is to bring a more systematic mutation screen for 2A peptides

2) The analysis of the P19X/G18X-induced ribosome stall: Examining the striking finding of a previous study not only offers insight of the event at the peptidyl-transferase centre, but also give a perspective to the further structural studies on the 2A peptides.

Chapter 2: Materials and Methods

2.1 materials

In present study, reagents and chemicals, except otherwise claimed, were purchased from Sigma-Aldrich Chemical Co., BDH Chemicals, Fisher, or New England Biolabs. The common reagents, included regularly used buffers, are listed as followed:

50X TAE buffer (1 litre) 242 g Tris Base, 100ml 0.5M EDTA pH 8.0, 57.2ml Glacial acetic acid, water to 1 L;

TE buffer (10X): 10 mM Tris-HCL, 1 mM EDTA, pH 8.0;
TEG buffer: 100 mM Tris-HCL pH8; 2 mM EDTA; 20% glucose;
SDS/NaOH: 0.2 M NaOH, 1% (v/v) SDS;
TBS buffer (10X): 500 mM Tris.HCl, pH 7.4 and 1500 mM NaCl;
TBE buffer (10X): Tris Base,890 mM, Boric Acid,890 mM, EDTA (pH 8.0),20 Mm;
Protein loading dye: 4% w/v SDS, 20% v/v glycerol, 0.125 mM Tris.HCl pH 6.8, 0.004% w/v bromophenol blue and 100 mM DTT (added just before using);

2.2 Strains and nucleotides

2.2.1 Bacterial strains

In present study, bacterial strains were mainly used as tools to produce plasmids. The Table 1.1 details the *E. coli* strains used during this project.

2.2.2 Yeast strains

In this project, yeast strains were exploited to test the separation activity of 2A and 2A

mutant peptides, *in vitro* and *in vivo*.

Table 2.1 includes all yeast strains involved in present project.

`

Table 2.1 STRAINS

Strains	Genotype	Source
E. coli		
DH10B	*F- endA1 recA1 galE15 galK6 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC)λ-*	J D Brown
S. cerevisiae		
JDY4	*a his3-Δ200, trp-Δ99, leu2, ade2-101, ura3-Δ99, cir^o*	J D Brown
JDY5	α *his3-Δ200, trp-Δ99, leu2, ade2-101, ura3-Δ99, cir^o*	

2.2.3 Plasmids used in this project

The plasmids used in this project are listed in Table 2.2

2.2.4 Oligonucleotides used during present study

Oligonucleotides were purchased from Sigma-Aldrich Biotech, and re-suspended to 100 pmol/μl for usage. Their sequence information are summarized in Table .2.3

2.3 Microbiological techniques

2.3.1 Growth of the bacterial strains

The *E. coli* strain was cultured at 37℃, in Luria-Bertani medium (LB, same below (1% w/v Tryptone; 0.5% w/v Yeast extract; 1% Sodium Chloride; Plus 2 % w/v agar for plate)), adding 100 μM ampicillin to select the transformants.

2.3.2 Transformation of bacterial strains

For transformation, *E. coli* cells were firstly prepared using $CaCl_2$ method (section 2.3.1.1), followed by the operation described in section 2.3.1.2.

2.3.1.1 $CaCl_2$ method for preparation of competent cells

5 ml of LB culture of DH10-B was grown at 37℃ overnight, then diluted 200 times in 100 ml LB. This was re-incubated until the culture reached OD_{600} 0.4-0.5. The cells were incubated on ice for 10 min, then centrifuged at 5000 rpm for 5 min. The resulting pellets were re-suspended at cold 40 ml TfbI and incubated on ice for 5 min. The cells were harvested at 3000 rpm for 10 min, then re-suspended in 4 ml TfbII. These re-suspended cells were incubated for 15 min on ice, then pipetted into 100 or 200 μl aliquots and stored at -80℃.

TfbI: 30mM KOAc, 100mM $RuCl_2$, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% Glycerol (adjusted to pH5.8 with Acetic Acid and filter sterilised).

TfbII: 10mM MOPS, 75mM $CaCl_2$, 10mM $RuCl_2$, 15% Glycerol (adjusted to pH6.5 with KOH and filter sterilised).

2.3.1.2 Transformation of competent cells

For transforming DNA into the competent cells, cells were thawed on ice, then mixed with DNA (each 50 μl cells with 1 μl plasmid DNA or 5 μl of ligation mix). The mixture was incubated on ice for 15-30 min, heat shocked at 42℃ for 1 min and returned to ice for 2 min. 1 ml of LB broth was added to the mixture, then incubated at 37℃ for 1 hour. The cells were then harvested at 2500 rpm for 2 min, the majority of the media was removed, cells were then re-suspended in the remaining solution and

spread on a LB ampicillin plate to select for the transformants.

2.3.3 Growth of the yeast strains

All yeast strains involved in this study were grown at 30℃, they were either propagated on rich YPD media (1% bacto-yeast extract, 2% peptone, 2% glucose Plus 2 % w/v agar for plate), or selective medium for *in vivo* test of the separation activity of 2A sequence/mutants, and selection for transformant.

2.3.4 Transformation of yeast strains

In this study, the plasmids were transformed into yeast cells using the lithium acetate method. For transformation, yeast cells were grown in 50 ml at 30℃ overnight and diluted 20 times, then re-incubated until the culture reached OD_{600} = 0.3-0.5. Cells were collected at 3000 rpm for 5 min, then re-suspended in 20 ml lithium acetate mix (0.1M LiOAc pH 7.3, 1XTE) and centrifuged again. The resulting pellet was re-suspended in 1 ml same mix, then each aliquot of the re-suspended cells was mixed with 0.1-1 µg DNA, 10 µl of 10mg/ml salmon-sperm DNA (previously heated at 100℃ for 5 min and returned into ice), and 700 µl of PEG mix (46%(w/v)polyethylene glycol-2000, 0.1M LiOAc pH7.3, 1XTE). The mixture was incubated at 30℃ for 1 hour then heat shocked at 42℃ for 15 min. The cells were pelleted by centrifuge at 5000 rpm for 1 min, then re-suspended in 1 ml YPD, and spread on desired selective plate for selection of the yeast transformants.

2.4 Molecular biology method

2.4.1 Phenol chloroform isoamylalcohol extraction/isopropanol precipitation

In present study, phenol chloroform isoamylalcohol (PCI) extraction was used to isolate and/or concentrate DNA and RNA from aqueous solution. The target solution was mixed with the equal volume of phenol chloroform isoamylalcohol (optimized pH for DNA: 5.5; for RNA: 4.5) thoroughly by vortex for 5 min. The mixture was centrifuged at 13000 rpm for 3 min, and the upper aqueous phase containing the nucleic acid was taken. This extraction might repeat several times, until the surface of two phases got clean. The resulting solution from extraction was added into equal volume of isopropanol and 1/10 volume of 3M KOAc (pH 5.3), centrifuged at 13000 rpm for 15 min. The pellet was washed with 70% ethanol, then resuspended in desired volume of TE buffer or water.

2.4.2 Plasmid preparation from E. coli

2.4.2.1 Mini-prep for 5 ml of culture

The mini-prep of plasmids was maily performed using QIAprep spin kit (Qiagen) and Wizard plus SV DNA purification system (Promega), as protocol provided by the manufacturer. Besides, small scale DNA preparations were also carried out by alkakine lysis method (described in section 2.4.2.2). In this case, the basic operation was same as large scale, with certain times scaled-down.

2.4.2.2 Large scale preparation of plasmid DNA

In large scale, plasmids were prepared using alkakine lysis method. 50-100 ml of overnight culture was pelleted at 3500 rpm for 15 min, the cells pellet was then re-suspended thoroughly in 2 ml TEG buffer, then 4 ml SDS/NaOH solution was added and mixed gently by inversion. 2.5M NaAOc (pH 5.2) was added then again mixed by inversion. The contents were centrifuged as earlier. The supernatant was taken and mixed with equal volume of isopropanol, then centrifuged again. The resulting pellet

was washed by 70% ethanol, air-dried for 10 min to remove the residual ethanol. The washed pellets was re-suspended in 0.5 ml of TE buffer, incubated with 5 µl of 10 mg /ml RNase A at 37℃ for 30 min. The resulting solution was extracted with PCI 3 times and the aqueous phase was precipitated by adding 1/10 volume of 3M NaAOc and equal volum of cold isopropanol. This mixture was centrifuged at 13000 rpm, then the DNA pellet, after washed by 70% ethanol, was resuspended in 100µl of TE or water.

2.4.3 Polymerase chain reaction(PCR)

The polymerase chain reactions (PCR) were used in this study not only for amplification specific regions of DNA, but also as a tool to introduce the mutations into DNA sequence.

2.4.3.1 Reaction mixture

Standard PCR reaction were performed as follows:

Mixture (100µl total volume):

1 µl DNA template

1 µl oligonucleotide primer 1 (100 µM)

1 µl oligonucleotide primer 2 (100 µM)

10µl 10x PCR buffer (100mM Tris pH8.3, 500mM KCL, 20mM $MgCl_2$, 0.1% w/v gelatin)

10 µl 10x PCR dNTPs (2 mM each)

1 µl taq DNA polymerase

76 µl ddH_2O

2.4.3.2 Programme

The Standard Programme as follow:

60 sec at 95℃ (initial denaturing step)

20 sec at 95℃ (cyclic denaturing step)

30 sec at 53℃ (cyclic annealing step)

120 sec at 72℃ (initial extension step)

300 sec at 72℃ (final polishing step)

In practice, in term of generation of mutants, a touchdown programme (annealing temperature ranged from 65℃ to 50℃) was applied. During present study, all PCR reaction were performed in a Takara PCR Thermal Cycler.

2.4.3.3 PCR products purification

For isolation of amplified DNA from reaction mixture, purification was performed using Qiagen gel purification kit, as manufacturer's instruction, or with PCI extraction/isopropanol precipitation.

2.4.4 DNA digestion

A standard DNA digestion in present study was performed as follow:

2 µl DNA (0.1-2µg)
1 µl desired restriction enzyme(s)
2 µl digestion buffer (depends the enzyme(s) used)

ddH_2O to 20 µl

The mixture was incubated at 37℃ (or appropriate temperature depends of enzyme used) for 1 hour then run on 1-2% agarose gel containing 0.005% ethidium bromide in 1x TA buffer, to check the result.

2.4.5 Oligoes annealing protocol

Two complementary oligo were re-suspended in the annealing buffer, into 100 pmol. The mixture of oligos was then put into PCR tubes. The tubes were incubated on a thermal cycler and using the following program: 1) 95℃ for 2 minutes, 2) 25℃ for 30 minutes. The resulting annealing product was used in DNA ligations at 4℃.

The annealing buffer was: 100 mM Potassium Acetate, 30mM HEPES (pH 7.5).

2.4.6 DNA ligation

For ligation reactions, the digested vector and DNA fragment were isolated by using PCI extraction/ isopropanol precipitation, and follow mixture was assembled:

1 µl Vector (50-200ng)
2 µl DNA fragment (molar ratio to the vector was 5:1)
2 µl 10x T4 DNA ligase buffer
1 µl DNA ligase

ddH_2O to 20 µl

The mixture was incubated at 25℃ for 1 hour or 4℃ overnight, then taken 10µl for *E.coli* transformation.

2.5 Western blot techniques

The western blot was used in this study for the verification of the extra protein binds appeared in *in vitro* translation experiment. Therefore, samples were protein mixture generated by *in vitro* translation system, rather than extracted from cells.

2.5.1 SDS-polyacrylamide gel electrophoresis

Due to the small size of target protein, samples were separated using a 10% SDS-polyacrylamide Tris-Glycine gel. After 2 hours running at 120V, the gel was either stained and fixed with Coomassie blue for 2A activity assay, or used for transfer to electro-transmembrane for western bloting.

The composition of 10% SDS-polyacrylamide Tris-Glycine gel:

1) Resolving gel (10 ml)

2.5 ml resolving gel buffer (1.5M Tris.HCl pH 8.8)
2.5 ml 40% w/v acrylamide
1.5 ml 2% w/v bis-acrylamide
0.1 ml 10% w/v SDS
3.25 ml H2O
4 µl TEMED
167 µl 10% w/v ammonium persulphate

2) Stacking gel (10 ml):

1.25 ml stacking gel buffer (1M Tris.HCl pH 6.8)
1.25 ml 40% w.v. acrylamide
0.66 ml 2% w/v bis-acrylamide
0.1 ml 10% w/v SDS
6.4 ml H2O
5 µl TEMED
120 µl 10% w/v ammonium persulphate

2.5.2 Transmembrane and blot

For western bloting, the protein separated by electrophoresis was transferred onto Protran nitrocellulose membrane (Schleicher and Schuell) using Biorad blotting apparatus at 20V for 1 hour. The membrane was stained by Ponceau for 3-4 min for visualization of transferred protein, then blocked by incubating in blotto (5% (w/v) non-fat milk, 0.5% (w/v) Tween 20, 1x TBS buffer), for 30 min. The membrane was then incubated with primary antibody (diluted 1000 times in blotto) at 4℃ overnight or room temperature for 2 hours. The blocked membrane was washed with blotto for 5 min three times, then incubating with secondary antibody (diluted 2000 times in blotto) for 1 hour. The membrane was washed 3 times in blotto (5 min each) followed by three times with TBS. The membrane was then incubated with ECL solution (12.5 mM luminol in DMSO, 0.2 mM p-coumaric acid, and 18% (w/v) H_2O_2 added before use), for 30 seconds to 2 min. X-ray film was used to detect the light emitted.

2.6 *in vitro* transcription/translation

In the present study, *in vitro* translations were used to test the separation activity of various 2A and 2A mutant sequences and generate the ribosome-peptide complex.

2.6.1 in vitro translation with coupled reticulocyte lysate

Coupled in vitro translation generated proteins from DNA substrate. In the project described in this thesis, coupled *in vitro* translation was performed with TNT T7 Coupled Reticulocyte Lysate System (promega). The operation was applied as manufacturer's recommended with several modifications as follows:

The follow reaction mix was assembled: (for 5µl reaction)

2.5µl Rabbit reticulocyte lysate (hand-warmed from -80℃)
0.2µl Reaction buffer
0.1µl T7 RNA polymerase
0.1µl Amino acid mixture, minus Methionine, 1mM
0.2µl [S^{35}]methionine, 1000Ci/mmol at 10 mCi/ml)
0.1µl RNasin Ribonuclease inhibitor, 40u/µl (not for generating the ribosome-peptide complex)
100 ng DNA template

Nuclease-free water to final volume of 5µl.

The mixture was then incubated at 30℃ for 1 hour (for activity assay) or 20 min (for production of ribosome-19X complex), then loaded onto a SDS-polyacrylamide gel or Nu-PAGE gel.

2.6.2 in vitro transcription

For generating mRNA from DNA substrate, *in vitro* transcription was applied with RiboMAX™ Large Scale RNAProduction System SP6 (promega). The follow reaction was assembled as protocol:

4 µl SP6 transcription 5X buffer

4 µl rNTPs (25 mM ATP, CTP, UTP and 3 mM GTP)

1-2 µg DNA template (linearized by restriction enzyme and treated with PCI extraction/ isopropanol precipitation)

1.5 µl Ribo m^7G cap analog, 40 mM (purchased from promega)

2 µl Enzyme mix (SP6)

Nuclease-free water to final volume of 20 µl.

(The cap-analog here was for generated the 5'-end capped mRNA that was competent for *in vitro* translation.)

The mixture was incubated at 37 ℃ for 5 hours, then treated with PIC extraction/isopropanol precipitation. The production of mRNA could be used directly for *in vitro* translation or stored at -80℃.

2.6.3 in vitro translation with reticulocyte lysate

Uncoupled *in vitro* translation was to produce generate protein from mRNA substrate. In this study, 3 kinds of different system were used for different aims.

The reticulocyte lysate used was Flexi® Rabbit Reticulocyte Lysate System (promega). The follow mixture was assembled according the protocol:

3.3µl Flexi Rabbit reticulocyte lysate

0.1µl Amino acid mixture, minus Methionine, 1mM

0.2µl [S^{35}]methionine, 1000Ci/mmol at 10 mCi/ml)

0.14µl KCl, 2.5M

500 ng RNA substrate (generated from 2.6.2)

Nuclease-free water to final volume of 5µl.

The mixture was incubated at 30℃ for 15-20 min, then ran on the Nu-PAGE gel immediately as described in section 2.6.6

2.6.4 in vitro translation with wheat germ lysate

The wheat germ lysate used in this study was Flexi® Rabbit Reticulocyte Lysate System (promega). The follow mixture was assembled according the protocol:

2.5µl Wheat germ extract

0.4µl Amino acid mixture, minus Methionine, 1mM

0.1µl [S^{35}]methionine, 1000Ci/mmol at 10 mCi/ml)

500 ng RNA substrate (generated from 2.6.2)

Nuclease-free water to final volume of 5µl.

The mixture was incubated at 30℃ for 15-20 min, then run on the Nu-PAGE gel immediately as described in section 2.6.6

2.6.5 in vitro translation with yeast extract

The yeast lysate used here was laboratory-made from strain JDY5. For *in vitro* translation, the mixture was assembled as follow:

1.2µl KOAc, 1M

2.3µl Reaction buffer (25 ml '6X buffer' plus 3µl creatine phosphokinase, and 7µl [S^{35}]methionine)

3.3µl Yeast lysate

500 ng RNA substrate (generated from 2.6.2)

Nuclease-free water to final volume of 10µl.

The mixture was incubated at 22℃ for 15-20 min, then ran on the Nu-PAGE gel immediately as described in section 2.6.6

'6X buffer':

66µl 1M HEPES-KOH pH 7.4,

6µl 1M $Mg(oAc)_2$,

93.8µl 800 mM creatine phosphate,

22.5µl 100 mM ATP,

3µl 100 mM GTP,

120µl Amino acid mixture, minus Methionine, 1mM

5.1µl 1M DDT,

183.6µl nuclease-free water

2.6.6 Nu-PAGE gel

For detection of ribosome-peptide complex, the Nu-PAGE gel (Invitrogen) was used to separate the samples synthesized by *in vitro* translation.

In this project, the samples were mixed with protein loading dye as described in 2.1, with adjustment of pH to 8.5 using NaOH. The samples were then run on the gel at 195 V for 30-40 min in 1X MES SDS running buffer. The gel was fixed in the fixing solution for at least 1 hour. The radioactivity signal of protein containing [S^{35}]methionine were detected by follow protocol described in section 2.6.7.

20 X MES SDS Running Buffer: (for 40 ml)

9.76 g	MES
6.06 g	Tris Base
1 g	SDS
0.3 g	EDTA
40 ml	nuclease-free water.

Fixing solution: (for 550 ml)

180 ml ethanol
100 ml formaldehyde
10-20 ml of glycerol (otherwise gel cracks during drying)
240 ml deionised water

2.6.7 Exposure to phosphor screen and analysis

In the study described in this thesis, the purse of radioactively labeled protein was detected by phosphor screen.

After running, the SDS-polyacrylamide gels were fixed with coomassie blue for 2 hours, then treated with DESTIN solution for 30 min; for Nu-PAGE gel, the gel was fixed with fixing solution for 1 hour.

The fixed gel was then dried with gel drier (Labconco) at 80℃ for 2 hours, and blocked with a phosphor image a phosphor image (BAS-MP 2040S) in a cassette, exposure for 2 hours to 2 days. The image was screen on the Typhoon Tiro scanner to detect the signal.

2.6.8 Inhibition of protease

For inhibition of the potential protein degradation during the *in vitro* translation, two protease inhibitors were tested in this study:

1) phenylmethylsulfonyl fluoride (PMSF): was made up to 100 mM in isopropanol and then added into the reaction mixture with a final concentration of 0.5 or 1 mM;
2) complete EDTA-free (Roche) mini protease inhibitor tablets (CMEF) : was added into reaction with a proportion of 1 tablets to 50 ml reaction mixture.

2.7 Purification of ribosome-peptide complex (RNC)

In this project, the method of RNC (with a polyhis-tag) purification from in vitro translation mixture was based with the one described in (Halic *et al*., 2004) with several modifications.

After *in vitro* translation described in 2.6, 40µl fresh samples (8X scale up as described in 2.6) was terminated with 2µl of 2 mg/ml cycloheximide, then carefully layered on a 600µl high-salt sucrose cushion in a centrifuge tube (Beckman Polycarbonate 11x34 mm). The tube was gently placed into an Optima TLX Ultracentrifuge (Beckman), spun

at 100000 rpm for 2 hours. The resulting pellet was resuspended in 40µl 250 buffer, (this step need 4 hours to overnight) then mixed with magnetic beads from 100µl rinse. The resulting mixture was incubated on a rotary sharker for 1 hour at 4℃, then the beads were collected and the supernatant was removed. The beads were washed with 0.5 ml 250 buffer and then 0.2 ml 500 buffer. The beads were then washed with 10 mM imidazole solution. The RNC was eluted by washing with 100 mM imidazole solution 2 times. The elutions containing RNC were collected, and analyzed on a Nu-PAGE gel as described in 2.6.6.

High-salt sucrose cushion: 50mM Tris-HCl (pH 7.0), 500mM potassium acetate, 25mM magnesium acetate, 2mM dithiothreitol (DTT), 1M sucrose and 10µg/ml cycloheximide added immediately before use

250 buffer: 50mM Tris-HCl (pH 7.0), 250mM potassium acetate, 25mM magnesium acetate, 0.1% (w/v) Nikkol, 5mM β-mercaptoethanol, 250mM sucrose, and 10µg/ml cycloheximide added immediately before use

500 buffer: 250 buffer containing 500mM potassium acetate

10 mM imidazole solution: 250 buffer containing 10 mM imidazole

100 mM imidazole solution: 250 buffer containing 100 mM imidazole

Table 2.2 PLASMIDS

Plasmid	Insert	Vector	Source
pJNY58	Ub-R-2A-ADE2	pRS314 *TRP1, CEN6, ARS4*	J D Brown (lab)
pJNY57	Ub-R-2A^{P17A}-ADE2		
pPS30	Ub-R-2A^{E14Q}-ADE2	pJNY58 *TRP1, CEN6, ARS4*	P Sharma
pPS31	Ub-R-2A^{S15I}-ADE2		
pPS33	Ub-R-2A^{N16H}-ADE2		
pPS34	Ub-R-2A^{N16E}-ADE2		
pPS35	Ub-R-2A^{N16Q}-ADE2		
pFYY1	Ub-R-2A^{P19A}-ADE2		This study
pFYY2	Ub-R-2A^{P19E}-ADE2		
pFYY3	Ub-R-2A^{P19G}-ADE2		
pFYY4	Ub-R-2A^{P19K}-ADE2		
pFYY5	Ub-R-2A^{P19R}-ADE2		
pFYY6	Ub-R-2A^{S15A}-ADE2		
pFYY7	Ub-R-2A^{S15E}-ADE2		
pFYY8	Ub-R-2A^{S15F}-ADE2		
pFYY9	Ub-R-2A^{S15G}-ADE2		
pFYY10	Ub-R-2A^{S15T}-ADE2		
pFYY11	Ub-R-2A^{S15P}-ADE2		
pFYY12	Ub-R-2A^{V13A}-ADE2		
pFYY13	Ub-R-2A^{V13E}-ADE2		
pFYY14	Ub-R-2A^{V13F}-ADE2		
pFYY15	Ub-R-2A^{V13G}-ADE2		
pFYY16	Ub-R-2A^{V13K}-ADE2		
pFYY17	Ub-R-2A^{V13L}-ADE2		
pFYY18	Ub-R-2A^{V13P}-ADE2		
pFYY19	Ub-R-2A^{G11A}-ADE2		
pFYY20	Ub-R-2A^{G11H}-ADE2		
pFYY21	Ub-R-2A^{G11K}-ADE2		
pFYY22	Ub-R-2A^{G11P}-ADE2		
pFYY23	Ub-R-2A^{L1V}-ADE2		
pFYY24	Ub-R-2A^{L6F}-ADE2		
pFYY25	Ub-R-2A^{L2W}-ADE2		
pFYY26	Ub-R-2A^{F4L}-ADE2		
pFYY27	Ub-R-2A^{L1G}-ADE2		
pFYY28	Ub-R-2A^{L2G}-ADE2		
pFYY29	Ub-R-2A^{N3G}-ADE2		
pFYY30	Ub-R-2A^{F4G}-ADE2		
pFYY31	Ub-R-2A^{D5G}-ADE2		
pFYY32	Ub-R-2A^{L6G}-ADE2		

pFYY33	Ub-R-2A^{L7G}-ADE2	pJNY58 *TRP1, CEN6, ARS4*	This study
pFYY34	Ub-R-2A^{K8G}-ADE2		
pFYY35	Ub-R-2A^{L9G}-ADE2		
pFYY36	Ub-R-2A^{A10G}-ADE2		
pFYY37	Ub-R-2A^{D12G}-ADE2		
pFYY38	Ub-R-2A^{E14G}-ADE2		
pFYY39	Ub-R-2A^{N16G}-ADE2		
pFYY40	Ub-R-2A^{P17G}-ADE2		
pFYY41	Ub-R-2A^{L1P}-ADE2		
pFYY42	Ub-R-2A^{L2P}-ADE2		
pFYY43	Ub-R-2A^{F4P}-ADE2		
pFYY44	Ub-R-2A^{L7P}-ADE2		
pFYY45	Ub-R-2A^{A10P}-ADE2		
pFYY46	Ub-R-2A^{G19P}-ADE2		
pFYY47	Ub-R-2A$^{L2W+F4L}$-ADE2		
pFYY48	Ub-R-2A$^{L1V+L6F}$-ADE2		
pLH135	GFP-2A-PAC	*In vitro* translation Using T7 promotor	R Martin (lab)
pLH136	GFP-2A^{P17A}-PAC		
pFY1	GFP-2A^{P17A}-PAC		This study
pFY2	GFP-2A^{E14Q}-PAC		
pFY3	GFP-2A^{S15I}-PAC		
pFY4	GFP-2A^{N16H}-PAC		
pFY5	GFP-2A^{N16E}-PAC		
pFY6	GFP-2A^{N16Q}-PAC		
pFY7	GFP-2A^{P19A}-PAC		
pFY8	GFP-2A^{P19K}-PAC		
pFY9	GFP-2A^{L1V}-PAC		
pFY10	GFP-2A^{L6F}-PAC		
pFY11	GFP-2A^{L2W}-PAC		
pFY12	GFP-2A^{F4L}-PAC		
pFY13	GFP-2A$^{L2W+F4L}$-PAC		
pFY14	GFP-2A$^{L1V+L6F}$-PAC		
pFY15	GFP-2A^{L1G}-PAC		
pFY16	GFP-2A^{L2G}-PAC		
pFY17	GFP-2A^{N3G}-PAC		
pFY18	GFP-2A^{F4G}-PAC		
pFY19	GFP-2A^{D5G}-PAC		
pFY20	GFP-2A^{L6G}-PAC		
pFY21	GFP-2A^{L7G}-PAC		
pFY22	GFP-2A^{K8G}-PAC		
pFY23	GFP-2A^{L9G}-PAC		
pFY24	GFP-2A^{A10G}-PAC		
pFY25	GFP-2A^{D12G}-PAC		
pFY26	GFP-2A^{V13G}-PAC		
pFY27	GFP-2A^{E14G}-PAC		

pFY28	GFP-2A^{S15G}-PAC	*In vitro* translation Using T7 promotor	This study
pFY29	GFP-2A^{N16G}-PAC		
pFY30	GFP-2A^{P17G}-PAC		
pFY31	GFP-2A^{P19G}-PAC		
pFY32	GFP-2A^{L1A}-PAC		
pFY33	GFP-2A^{L2A}-PAC		
pFY34	GFP-2A^{N3A}-PAC		
pFY35	GFP-2A^{F4A}-PAC		
pFY36	GFP-2A^{D5A}-PAC		
pFY37	GFP-2A^{L6A}-PAC		
pFY38	GFP-2A^{L7A}-PAC		
pFY39	GFP-2A^{K8A}-PAC		
pFY40	GFP-2A^{L9A}-PAC		
pFY41	GFP-2A^{G11A}-PAC		
pFY42	GFP-2A^{D12A}-PAC		
pFY43	GFP-2A^{V13A}-PAC		
pFY44	GFP-2A^{E14A}-PAC		
pFY45	GFP-2A^{S15A}-PAC		
pFY46	GFP-2A^{N16A}-PAC		
pFY47	GFP-2A^{P17A}-PAC		
pFY48	GFP-2A^{G18A}-PAC		
pFY49	GFP-2A^{P19A}-PAC		
pFY50	GFP-2A^{L1P}-PAC		
pFY51	GFP-2A^{L2P}-PAC		
pFY52	GFP-2A^{N3P}-PAC		
pFY53	GFP-2A^{F4P}-PAC		
pFY54	GFP-2A^{D5P}-PAC		
pFY55	GFP-2A^{L6P}-PAC		
pFY56	GFP-2A^{L7P}-PAC		
pFY57	GFP-2A^{K8P}-PAC		
pFY58	GFP-2A^{L9P}-PAC		
pFY59	GFP-2A^{A10P}-PAC		
pFY60	GFP-2A^{G11P}-PAC		
pFY61	GFP-2A^{D12P}-PAC		
pFY62	GFP-2A^{V13P}-PAC		
pFY63	GFP-2A^{E14P}-PAC		
pFY64	GFP-2A^{S15P}-PAC		
pFY65	GFP-2A^{N16P}-PAC		
pFY66	GFP-2A^{G18P}-PAC		
pFY67	GFP-2A^{TMEV}-PAC		
pFY68	GFP-2A^{EMCV}-PAC		
pFY69	GFP-2A^{10T+9F}-PAC		
pFY70	GFP-2A^{9T+10F}-PAC		
pFY71	GFP-2A^{9F+10T}-PAC		
pFY72	GFP-2A^{9F+10E}-PAC		

pFY73	GFP-2A^{9E+10F}-PAC	*In vitro* translation Using T7 promotor	This study
pFY74	GFP-2A^{A10I}-PAC		
pFY75	GFP-2A^{L9K}-PAC		
pFY76	GFP-2A^{L7Q}-PAC		
pFY77	GFP-2A^{L7D}-PAC		
pFY78	GFP-2A$^{A10I+G11H}$-PAC		
pFY79	GFP-2A$^{L9K+G11H}$-PAC		
pFY80	GFP-2A$^{L7Q+G11H}$-PAC		
pFY81	GFP-2A$^{L7D+G11H}$-PAC		
pFY83	GFP-2A^{KQ}-PAC[1]		
pFY84	GFP-2A^{YYKQ}-PAC[2]		
pFY85	GFP-2A^{P19X}-PAC		
pFY86	GFP-2A^{G18X}-PAC		
pFY87	GFP-2A^{P17X}-PAC		
pFY88	GFP-2A^{K8X}-PAC		
pFY89	GFP-2A$^{P17A+P19X}$-PAC		
pJN141	pro-α-factor-2A-prolactin	*In vitro* translation Using SP6 promotor	V Doronina
pJN275	Histag-pro-α-factor-2A-PLL		
pJN275M4	Histag-pro-α-factor-2A-PLL[3]		This study
pFY90	Histag-pro-α-factor-2A $^{TMEV(T15S)}$-PLL		
pFY91	Histag-pro-α-factor-2A^{IMNV}-PLL		
pFY92	Histag-pro-α-factor-2A $^{Prv-1}$-PLL		

1. ‘KQ’ refers to the FMDV 2A mutant L6K+L7Q+G11H
2. ‘YYKQ’ refers to the FMDV 2A mutant F4Y+D5Y+L6K+L7Q+G11H
3. See section 6.2.1 for detail

Table 2.3 OLIGOS

Oligo Name	Sequence (5'-3')	**Refer to Section**
Ubi5	CCAGGAATCCATGCAGATTTTCGTC	**Section 3.1**
AD3	GTGTAGGAACATCAACATGCTC	
135SEQ	CGGATCGACGGTGTGG	
135SEQ long	TGCCGATGTCGAGCC	
EMCVfor	CTAGAGGAGCATGCCACTACGCTGGTTACTTTGC GGACCTACTGATTCATGACATTGAGACAAATCCC GGGCC	**Section 5.2.1**
EMCVrev	CGGGATTTGTCTCAATGTCATGAATCAGTAGGTC CGCAAAGTAACCAGCGTAGTGGCATGCTCCT	
TMEVfor	CTAGAGGAGCATGCTACCATGCTGACTACTACAA ACAGAGACTCATACATGATGTAGAAATGAACCCC GGGCC	
TMEVrev	CGGGGTTCATTTCTACATCATGTATGAGTCTCTGT TTGTAGTAGTCAGCATGGTAGCATGCTCCT	
F+Efor	CTAGAGGAGCATGCCAGCTGTTGAATTTTGACCT TCTTAAGCTTATTCATGACATTGAGACAAATCCCG GGCC	**Section 5.2.2**
F+Erev	CGGGATTTGTCTCAATGTCATGAATAAGCTTAAG AAGGTCAAAATTCAACAGCTGGCATGCTCCT	
F+Tfor	CTAGAGGAGCATGCCAGCTGTTGAATTTTGACCT TCTTAAGCTTGCGCATGATGTTGAAATGAATCCTG GGCC	
F+Trev	CAGGATTCATTTCAACATCATGCGCAAGCTTAAG AAGGTCAAAATTCAACAGCTGGCATGCTCCT	
T+Ffor	CTAGAGGAGCATGCTACCATGCTGACTACTACAA ACAGAGACTCGCGGGAGACGTCGAGTCCAACCC TGGG-CC	
T+Frev	CAGGGTTGGACTCGACGTCTCCCGCGAGTCTCTG TTTGTAGTAGTCAGCATGGTAGCATGCTCCT	
E+Ffor	CTAGAGGAGCATGCCACTACGCTGGTTACTTTGC GGACCTACTGGCGGGAGACGTCGAGTCCAACCC TGGGCC	
E+Frev	CAGGGTTGGACTCGACGTCTCCCGCCAGTAGGTC CGCAAAGTAACCAGCGTAGTGGCATGCTCCT	
T+F2for	CTAGAGGAGCATGCTACCATGCTGACTACTACAA ACAGAGACTCATAGGAGACGTCGAGTCCAACCC TGGGCC	

T+F2rev	CAGGGTTGGACTCGACGTCTCCTATGAGTCTCTG TTTGTAGTAGTCAGCATGGTAGCATGCTCCT	**Section 5.2.2**
YYKQfor	CTAGAGGAGCATGCCAGCTGTTGAATTATTATAA ACAAAAGCTTGCGCACGACGTCGAGTCCAACCC TGGGCC	**Section 5.2.3**
YYKQrev	CAGGGTTGGACTCGACGTCGTGCGCAAGCTTTT GTTTATAATAATTCAACAGCTGGCATGCTCCT	
KQ-for	CTAGAGGAGCATGCCAGCTGTTGAATTTTGACAA ACAAAAGCTTGCGCACGACGTCGAGTCCAACCC TGGGCC	
KQ-rev	CAGGGTTGGACTCGACGTCGTGCGCAAGCTTTT GTTGTCAAAATTCAACAGCTGGCATGCTCCT	
T3	AATTAACCCTCACTAAAGGG	**Section 6.1.1**
T7	AATACGACTCACTATAGG	
SP6	CATTTAGGTGACACTATAG	
275for	AATTCAGATATCCATAGTTCAAACAAGAAGATT ACAAACTATCAATTTCATACACAATATAAACGG CCGCCACCATGGTA	**Section 6.2.1**
275rev	AATTTACCATGGTGGCGGCCGTTTATATTGTGTA TGAAATTGATAGTTTGTAATCTTCTTGTTTGAAC TATGGATATCTG	
SEQ141	AATGAACCCTTTGCC	
M4forTMEV	CTAGGGCATGCGCAGTACGAGGATACCACGCAG ACTACTACAAACAACGACTAATACACGACGTCGA GTCCAACCCTGGGCC	**Section 6.1.2**
M4revTMEV	CAGGGTTGGACTCGACGTCGTGTATTAGTCGTTG TTTGTAGTAGTCTGCGTGGTATCCTCGTACTGCGC ATGCC	
M4forPrv-1	CTAGGGCATGCAACTCAGGATACGTAGTAGGAGG ACGAGGATCACTACTAACATGCGGAGACGTCGA GTCCAACCCTGGGCC	
M4revPrv-1	CAGGGTTGGACTCGACGTCTCCGCATGTTAGTAG TGATCCTCGTCCTCCTACTACGTATCCTGAGTTGC ATGCC	
M4forIMNV	CTAGGGCATGCGAAATATCAGACTGCATGCTACC ACCACCAGACCTAACATCATGCGGAGACGTCGA GTCCAACCCTGGGCC	
M4revIMNV	CAGGGTTGGACTCGACGTCTCCGCATGATGTTAG GTCTGGTGGTGGTAGCATGCAGTCTGATATTTCG CATGCC	
Pur-17A	TTAGGCCCCGGGCCCGGCGTTGGACTCGACGTC TCC	**Section 6.2.2**

Pur-19x	TTAGGCCCCTTACCCAGGGTTGGACTCGACGTCTCC	**Section 6.2.2**
Pur-wt	TTAGGCCCCGGGCCCAGGGTTGGACTCGACGTCTCC	
Pur-19x17A	TTAGGCCCCTTACCCGGCGTTGGACTCGACGTCTCC	

Note: Due to the number and similarity, the oligoes used to generate FMDV 2A mutants from pJNY58 is not listed in table 2.3. See section 3.1 for detail.

Chapter 3: Site-specific mutagenesis analysis of 2A sequence

The project described in this thesis follows on from that of a former PhD student in the laboratory, Dr. Pamila Sharma. During her study, Dr. Sharma carried out a random mutagenesis of the FMDV 2A sequence to identify the key elements within it. The first part of the present study was a complementary site-specific mutagenesis study, to extend her finding.

This chapter summarizes the key findings of the previous screen, and the first part of present study. Before this however, an introduction is to give for the methods used for site-specific mutagenesis and the assays of 2A separation activity, which were the main strategy not only for the study described in this chapter, but also for that of chapters 4 and 5.

3.1 2A activity reporter systems

The activity of various 2A sequences and 2A mutants were reported by 2 independent systems: an *in vivo* assay that exploited the varied colour (white to red) of yeast expressing reducing amounts of Ade2p and an *in vitro* translation assay. The 2A used in these experiments was, with the exception of some experiments reported in Chapter 5, from FMDV.

3.1.1 In vivo reporter

The *in vivo* reporter system used in this study (Ub-R-2A/2A*-ADE2) was constructed by Jeremy Brown and published in (Sharma *et al*., 2012). This reporter comprises a fusion of sequences encoding ubiquitin, a single arginine, 2A or 2A mutants and Ade2p into a single open reading frame (UBI-R2A-ADE2). The vector that hosts this open reading frame is pRS314, a low copy (*CEN*) *S. cerevisiae-E.coli* shuttle plasmid containing the *TRP1* marker for selection in yeast (Sikorski *et al*., 1989). For the activity assay, this construct and variants of it containing different 2A mutants were transformed into *S. cerevisiae* (strain JDY4, see Table 2.1).

In *S. cerevisiae*, Ade2p is a key enzyme in synthesis of purine nucleotides converting phosphoribosylaminoimidazole (AIR) into phosphoribosylaminoimidazole carboxylate (CAIR), which is then converted into SAICAR by Ade1p.(Figure 3.1) Both AIR and CAIR are toxic to the cell and are oxidised and transported to the vacuole, where they accumulate as red-coloured derivatives. Therefore, cells lacking functional Ade2p or Ade1p grow as red colonies, particularly when adenine is limited in the media as this results in induction of the adenine biosynthesis pathway.

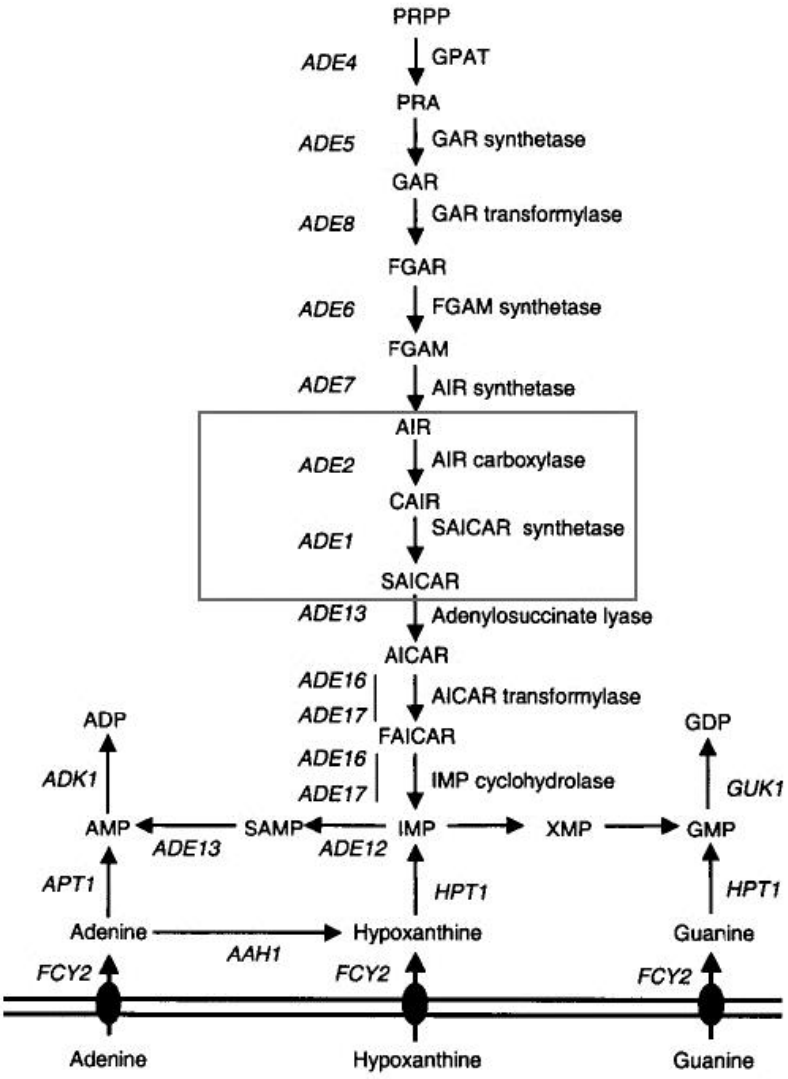


Figure 3.1: The Adenine Metabolic Pathway. Take notice on the crucial roles of *ADE1* and *ADE2* in transformation of AIR to SAICAR (Highlighted by gray rectangle). Taken from Rebora *et al*., (2001).

In cells transformed with the above-mentioned reporter where a mutation has been introduced into 2A that inactivates it, the translation product is the full length fusion of ubiquitin, arginine, 2A mutants and Ade2p. The N-terminal ubiquitin chain of this fusion polyprotien is proteolytically removed from the protein in the cytoplasm by ubiquitin-specific proteases and leaves arginine as a new N-terminus to the remainder of the protein. Proteins with an arginine at the N-terminus are very unstable and are rapidly degraded by the proteasome through the conserved N-end rule pathway (Varshavsky *et al*., 1997). Therefore, Ade2p is not stably expressed, and in otherwise, genomically *ade2* cells, such 2A mutant-containing reporters yield red or pink colonies on plates containing limiting amounts of adenine（see Figure 3.2a 'pRS314'~'E14Q').

On the other hand, *ade2* cells transformed with a plasmid containing wild type 2A or a 2A mutant with full or partial activity, the 2A reaction co-translationally separates the upstream Ub-R-2A from the downstream Ade2p. This results in stable expression of Ade2p and AIR and CAIR will be converted via Ade2p and Ade1p into SAICAR. When grown on adenine-limiting plates, this yields colonies with colours ranging from white through varying shades of pink depending on the level of activity of the 2A/2A mutant (see Figure 3.2a 'WT').

In conclusion, the fusion protein of ubiquitin, arginine, 2A or 2A mutants and ADE2p offers a reporter by which we can judge whether a 2A mutant is active. Moreover, the activity can be assessed by the relative colour development of colonies carrying different 2A/2A mutants.

An example of this is shown in Figure 3.2a, in which previously characterised 2A mutants (in Donnelly *et al*., (2001)) are compared, by spotting suspensions of cells containing UBI-R2A-ADE2 reporters with the various mutants onto a plate containing a limited amount of adenine. Here, as expected, cells containing the reporter with wild type 2A remained white, while those with the mutants E14Q, S15I or N16H that were

reported to have partial activities (56%, 42% and 31% respectively) conferred a pink colour on the cells, though, due to unknown reason, the shade of pink was slightly deeper with E14Q (56%) than with S15I (42%). In contrast, three mutants assessed previously with low or null activities, N16E (19%), N16Q (10%), P17A(0%) conferred a red colour to the cells transformed with the reporters carrying them. These results indicated, in general, that the colour phenotype of the cells transformed with various 2A mutants fit with the reported activity of these mutants. This set of mutants was used as control in followed study.

3.1.2 In vitro reporter

The polyprotein reporter used for the *in vitro* analysis was a gift from Professor Martin Ryan of St Andrews University. It comprises CFP (cyan fluorescent protein), followed by 2A or 2A mutant and PAC (puromycin N-acetyltranferase) which, when 2A or 2A mutant is active, is divided into upstream (CFP-2A) and downstream products (PAC). Both CFP and PAC are preceeded by a single copy of the V5 epitope tag. This construct was assembled in pcDNA3.1 (Invitrogen) and transcription is driven from a T7 promoter upstream of the open reading frame.

For testing the activity of 2A or 2A mutants, the above-mentioned plasmid containing the target 2A sequence was used in coupled *in vitro* transcription-translation reactions in reticulocyte lysate with [35S] methionine. (as described in 2.6.1) Products were then resolved on 10% SDS-polyacrylamide Tris-Glycine gels (2.5.1), which were fixed and dried, and then analyzed by phosphorimaging (2.6.7).

To quantify 2A activity, the incorporation of [35S] methionine into the translation

products [CFP-2A-PAC] (full length form), and the upstream- and downstream-products PAC and [CFP-2A] was determined by phosphorimaging. The photo-stimulated luminescence (PSL) of each band was determined, background subtracted and then divided by the methionine content (CFP:10, PAC:10) of the appropriate translation product (PSL^{corr}). 2A activity (%) was calculated as:

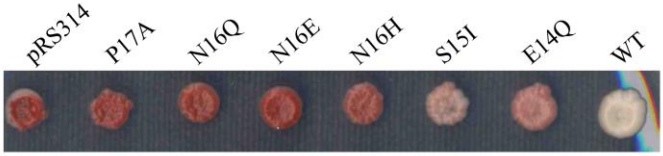
$$[PSL^{corr} \text{ of CFP2A}] / ([PSL^{corr} \text{ of CFP2APAC}] + [PSL^{corr} \text{ of CFP-2A}])$$

As the 2A reaction is co-translational, here we used the ratio of N-terminal product (CFP-2A) to full length plus N-terminal product to determine the 2A efficiencies, namely the percentage of ribosomes that released N-terminal product. It is notable that, the above formula is different to the one used in previous study, (Donnelly *et al*., 2001) in which the 2A activity is determined by the ratio of the total of separated products (up- and down-stream) to the sum of all products (up- and down-stream products plus the full length protein). This formula measures the percentage of translation products resulted by the 2A reaction.

An example *in vitro* assay is shown in Figure 3.2b, with the same mutants set as 3.1.1. The quantified data (table 3.1) shared the comparable pattern as the results obtained previously, with an exception on mutants E14Q and S15I, of which in the previous study the former was significantly higher than the latter, (56% and 42%) but were very similar in present assay. The difference of exact value between present and previous study can be attributed mainly to the alteration of calculation equation. Indeed, when we quantified the present in vitro assay with the formula used in (Donnelly *et al*., 2001), more similar values were observed (Compare the columns 2 and 4 of table 3.1).

The present data were also comparable with *in vivo* test, except the mutants E14Q and S15I, where the *in vivo* assay scored E14Q as less active than S15I, but *in vitro* assay revealed them to be similar, as Figure 3.2 and table 3.1 shown.

a.Example of *in vivo* assay



b.Example of *in vitro* assay

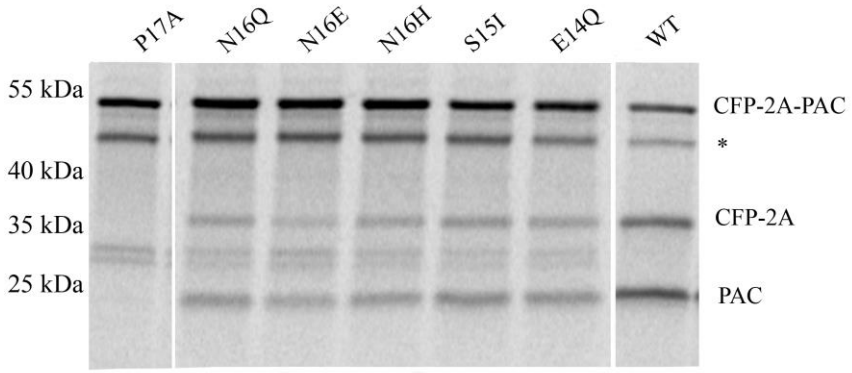


Figure 3.2 *in vivo* and *in vitro* assays of 2A activity on wild type 2A and 2A mutants P17A, N16Q, N16E, N16H, S15I and E14Q.

2A mutants	Present assay	Previous reported	Present assay previous cal.
WT	70.8 %	~90 %	81.7 %
E14Q	18.5 %	56 %	32.2 %
S15I	18.1 %	42 %	32.6 %
N16H	9.9 %	31 %	16.4 %
N16E	6.8 %	19 %	13.1 %
N16Q	6.2 %	10 %	7.2 %

Table 3.1 Quantification data in the previous study (Donnelly *et al*., 2001) (column 2 'Previous reported'), the present assay (column 1 'Present assay'), and the present data analysed by formula used in (Donnelly *et al*., 2001) (column 3 'Present assay previous cal.')

3.1.2.1 A discussion of extra bands on the in vitro test

SDS-PAGE gels of CFP-2A-PAC translation products revealed an 'extra' band at about 50 kDa just below the full length translation product (indicated with * in Figure 3.2b). Attempts were made to characterise this band as, if it was derived from CFP-2A-PAC taking it into account in calculating 2A activity may be important.

The 50 kDa band was also observed in experiments performed in Prof. Ryan's lab. (personal communication). Initially it was considered that it may be a degradation product from the full length protein. Therefore, protease inhibitors phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor tablets (Roche, Complete, Mini, EDTA-free were added into coupled transcription/translation reactions, and

reaction times were reduced, in order to reduce any proteolysis (protocol see section 2.6.8). However, as shown in Figure 3.3a, these measures did not affect the outcome. Thus the 50 kDa species may not be the production of degradation.

To confirm whether or not the extra band was a part of construct [CFP-2A*-PAC], the products of coupled *in vitro* transcription translation were resolved on SDS-PAGE gels, transferred to nitrocellulose and probed with anti-V5 monoclonal antibodies as described in 2.5. As both CFP and PAC portions of the polyprotein contain a V5 epitops, the separated and full length translation product should all be detected by the antibody. Following Western blotting, phosphorimaging was used to detect the [35S] incorporated into the poteins during translation. The outcomes of western blot and phosphorimaging are shown in Figure 3.3b. The results of the two detection methods superimposed to each other: the 50 kDa band was detected in both cases, indicating that it must be a truncated version of the full length (i.e. unresolved) translation product. The mechanism through which this occurs still lacks explanation. One possibility is that during translation initiation, the minority of the initiation complexes 'skip' the first methionine and initiate translation at a second methionine, but no experiment was carried out to confirm this. However, a full length translation product generated from the second methionine in the protein has a predicted molecular weight of 45 kDa, which could be consistent with this possiility. Intriguingly, we found that the 50 kDa band only appeared while using the *in vitro* translation system based on reticulocyte lysate. When translation preformed with the same pLH135 construct but in wheat germ this product was not generated (figure 3.3c). Although the mechanism by which the 50 kDa band is generated remains unclear, it was decided to use reticulocyte lysate to assay the 2A activity *in vitro*, for the following reasons:

1 The key goal of this study was to compare activity of different mutants, the loss of small amount of full length protein does not affect the *relative* activities of mutants.

2 When the efficiency of 2A peptides was compared between coupled transcription-translation and direct translation of RNA of wheat germ, similar values were obtained.

3 With DNA as the template, the coupled system is significantly simpler to use and is more economic than generating RNA and then using it as a template for translation. This was particularly important given the large number of mutants involved in this study. The only coupled system available, however, is the reticulocyte lysate.

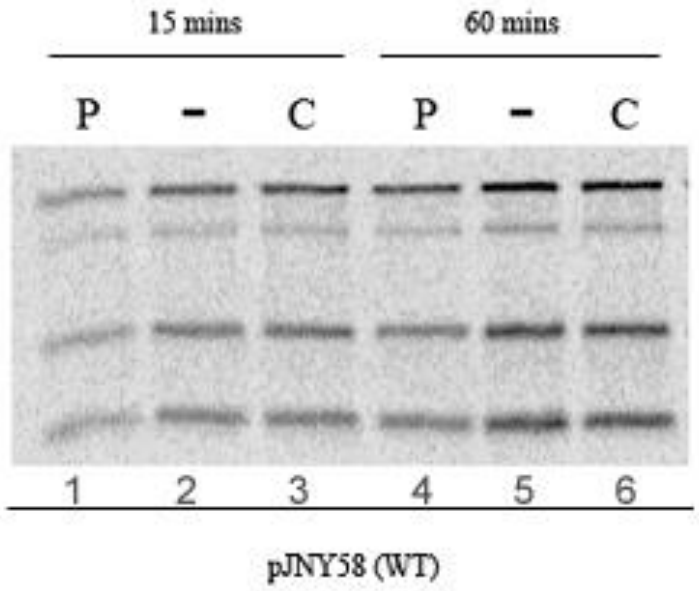


Figure 3.3a *in vitro* translation on wild type 2A with protease inhibitors treatment. The PMSF (lane 1,4) or protease inhibitor tablets (CMEF) (lane 3, 6), or water (lane 2,5) were added to the *in vitro* assay mix before the reaction, as protocol. The reactions were implemented for 15mins (lane 1~3) or 60mins (lane 4~6).

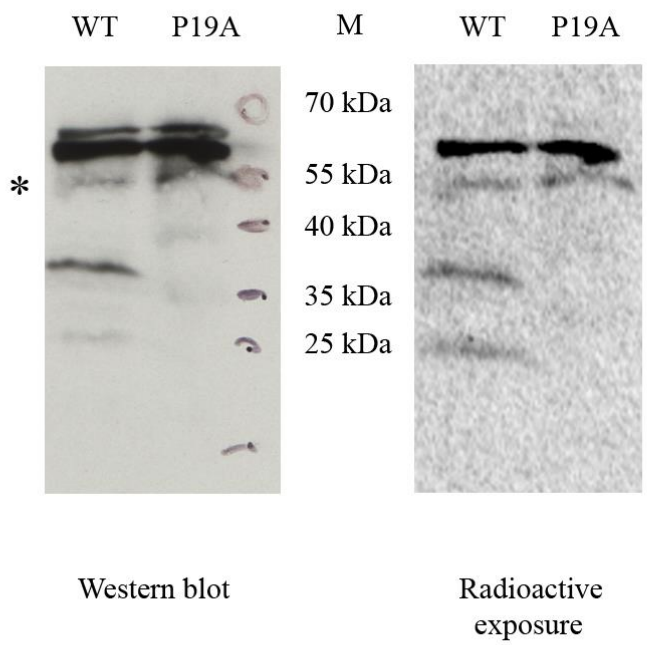


Figure 3.3b Western blot and western blot and phosphorimaging of wild type 2A and 2A mutant P19A (as an inactive control). See text for detail. '*' marks the 50 kDa translation product discussed in the text and which may be caused by leaky scanning and initiation at the second methionine in the protein.

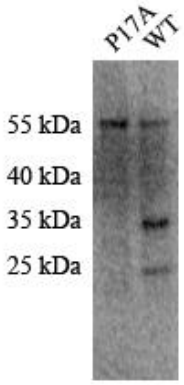


Figure 3.3c *in vitro* translation in wheat germ, on the wild type 2A and inactive mutant P17A. See text for detail

3.1.3 Generation of 2A mutants

To generate mutations in 2A of FMDV, PCR was used to amplify fragments containing the desired mutations from plasmid pJNY58, which contained the Ub-R-2A-ADE2 reporter (the strategy is outlined in Figure 3.4).

Primers were used specifying the desired changes to the 2A coding sequence. These primers were designed as reverse primers and with an NcoI site at their 5'-end. The forward primer used in the amplification is a common primer 'Ubi5' which contained a BamHI restriction site at the 5'-end. A touchdown PCR programme (annealing temperature set at 65 ℃, and reduced during the reaction to 50 ℃) was used to amplify the PCR fragments (see section 2.4.3). The resulting PCR fragments were then cloned into plasmids pJNY58 (for *in vivo* assay) and pLH135 (for *in vitro* study) using the BamHI and NcoI sites, and internal XbaI and ApaI sites respectively, replacing the wild type 2A coding sequence with the mutant (Figure 3.4). All plasmids were sequenced to confirm both the presence of the desired mutations within the 2A coding sequence, and that no other mutations had been introduced during PCR.

The mutants with amino acids replacement in position 1~17th were generated as described above. However, it was noted that, the two final condons of 2A, Gly18 and Pro19, is corresponding to the 'gggccc' in the DNA sequence. This formed an ApaI site that was to be manipulated in following step of the construct of the *in vitro* reporter pLH135. Hence, the mutants related to these two positions were added an extra ApaI site 2 condons downstream to the 2A coding region, during design of the primers used to introduce the mutations.

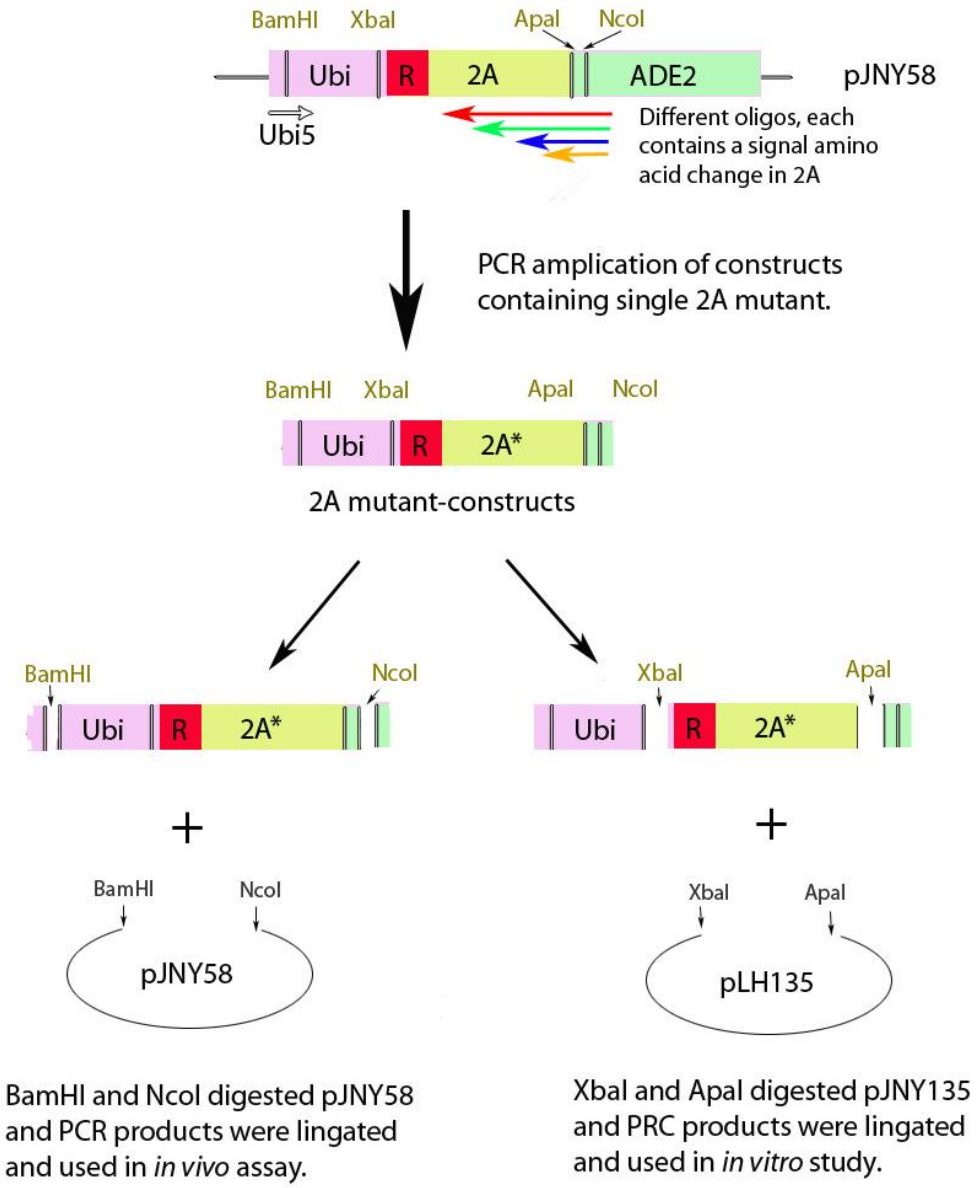


Figure 3.4: The general cloning strategy for *in vivo* and *in vitro* studies, see text for details.

3.2 Summary of previous random screen

A random mutagenesis had been carried out in the laboratory by Dr. Pamila Sharma prior to my studies. This was to generate a broad range of mutations in 2A, and added significantly to the mutants that had been generated by site-directed mutagenesis previously (Donnelly *et al*., 1997; Donnelly *et al*., 2001). In these studies, carried out *in vitro*, several important features of 2A sequence were revealed (see introduction section 1.3.2 for details) However, the number of the mutants tested was relatively small. In addition, these mutations were largely within the C-terminal conserved motif of 2A, and the role of N-terminal section of the peptide was not explored. To overcome these limitations and provide a more complete picture of the effects of mutation of sequence on the function of the 2A peptide, Dr Pamila Sharma carried out a large scale random screen on 2A from FMDV. In that study, random mutations were introduced into the C-terminal portion (Asp14 to Pro19) of 2A by 2-step PCR with a primer containing 2% non-template nucleotide in the region corresponding to Asp14 to Pro19. *In vivo* gap repair (Muhlrad *et al*., 1992) incorporated the mutated sequences into Ub-R-2A-ADE2, and growth on plates containing limited amounts of adenine allowed identification of the relative activity of the mutants. A similar strategy was also applied to N-terminal section of 2A (Leu1-Gly11). The activity of these mutant 2A was analysed using the *in vivo* reporter system (as described in section 3.1). 400 yeast colonies containing 2A mutants, with different colour (white-2A not affected by mutation or Pink-2A affected by mutation) were selected from initial plates and the 2A coding region of the plasmids they contained was amplified and sequenced to identify mutations within it.

From 400 colonies, 354 2A sequences were clearly identified. Excluding the ones without mutation or with the same mutations, 56 independent single mutants and 29 double mutants were identified. These mutants and their 2A activity (reported by the colour of their host colonies) are summarize in table 3.2

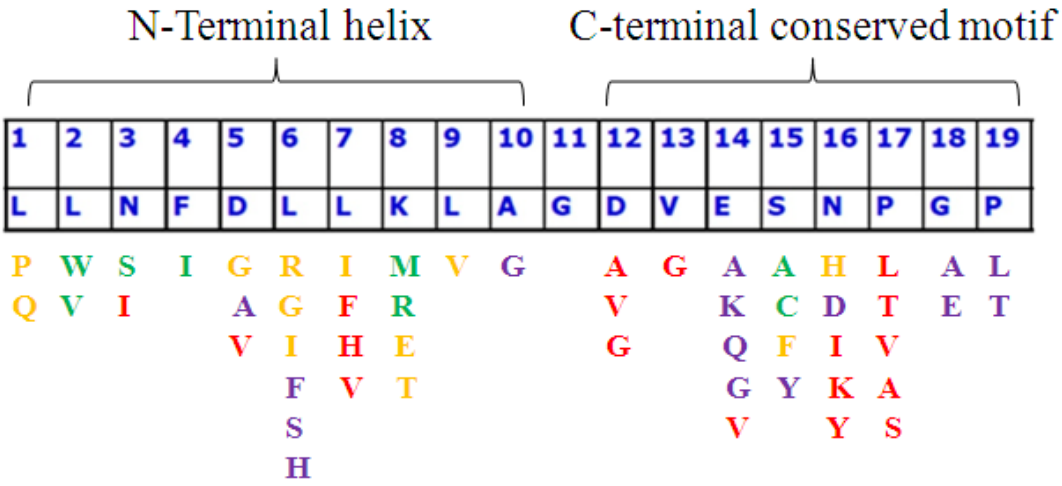


Table 3.2 Summary of single mutants identified in previous random screen. Data are summarised from Dr Sharma's previous *in vivo* experiments. The 2A activity was roughly estimated by comparison with several control mutants (E14Q, S15I, N16H, N16E, and N16Q) The letters in the form represent mutation i.e. the corresponding amino acid replacement in the position; the colour of the letter represent the 2A activity of the mutation estimated by comparison to the control:

Red refers the mutant is inactive; and the JDY4 colony give red colour; Purple refers the mutant having weak activity; and the JDY4 colony give deep pink colour;

Yellow refers the mutant having significant activity; and the JDY4 colony gives light pink colour;green refers the mutant having an activity close to wild type 2A; and the JDY4 colony give white colour.

The most important conclusion from this study is that the 2A sequence, even the non-conserved portion, is largely intolerant to change. Among 200 white colonies (i.e. with full 2A activity) analysed by Dr. Sharma, only 8 contained changes to the peptide sequence. Moreover, as can be seen in table 3.1, although in general mutations to the conserved portion impaired activity to a greater degree, several single alterations within the non-conserved N-terminal part also decreased 2A function dramatically (for example, mutant N3I only retains a small amount of activity) This was in some degree surprising, as the previous alignment study revealed that the N-terminal portion is highly variable in different 2A sequences. These results indicated that the importance and function of the N-terminal portion of the peptide needs re-evaluation.

Another finding from the random screen was that two of the double mutants identified (L1V+L6F and L2W+F4L) had less reduction of 2A activity than mutants containing only one of the amino acid changes. In this study, the mutant L2W+F4L conferred white colony colour corresponding to a full 2A activity, however, F4L displayed only partial activity (pink colony) in the *in vivo* assay. Similarly, cells transformed with the reporter containing the L6F mutation developed a darker pink colour than those containing the L1V+L6F 2A variant, indicating a higher activity in the double mutant. As in both cases, the other single mutant (L2W and L1V) was not isolated and the activity of them is unknown, the basis of the relatively high activity of the double mutants is not known. One explanation of this result is that 2 mutants might complement to each other, which is interesting and, if confirmed, might be very helpful in extending our understanding of the 2A reaction mechanism.

3.3 Generation of further mutations within 2A

While the random mutagenesis screen was very revealing, it had some limitations. The most significant of these was that, at several positions, few or no mutants were isolated. For filling the 'holes' left as well as expanding the finding of the mutagenesis screen, a number of further single mutants were generated using a site-specific mutagenesis strategy, and examined *in vivo* and/or *in vitro*. The main limitations of the random screen and the mutants designed in the present study to supplement them are listed as below:

While the 2 mutations identified at the final position 19 had very little activity, (P19T and P19L, see Table 3.2) consistent with the previous study suggesting Pro19 crucial for the function of 2A sequence, the limited number of mutations to the 19th position limited the strength of this argument. Therefore, a larger collection of amino acid, comprising A,E,G,K,R, were introduced at 19th position to replace the proline, to test whether or not other amino acids can function at this position.

Alterations to Ser15 showed relatively high activity, with the exception of substitution with Tyr. As position 15 of 2A is one of only 2 within the C-terminal region that is variable, these data suggest that this is not an important amino acid in terms of mechanism. For determining whether a broad or limited range of amino acids at this position is compatible with full 2A function, here Ser15 was further replaced with A,E,F,G,T,P and I at this position.

The other position that is variable within the C-terminal portion of 2A is amino acid 13, valine in the FMDV peptide, though this only extends to Valine and Isoleucine. No mutations were recovered at this position in the random screen, and similarly no changes to Gly11 were isolated. To generate a more complete picture of the effects of mutation of 2A sequence on its efficiency, in the present study, a set of amino acids were substituted for Gly or Val at these positions (A,E,F,G,K,L,P to Val13 and A, H, K, P to Gly11).

In addition to the mutants above, several were made specifically to test the possibility that double mutants isolated in the screen had more activity than single mutants from which they were composed.

3.3.1 in vivo test

The above mutants were inserted into the Ubi-R-2A-ADE2 reporter and transformed into the yeast strain Y4 for *in vivo* assay, as described in section 3.1.3 and 3.1.1. After selection for the plasmids on media lacking tryptophan, colonies of each mutant were isolated and spotted onto plates containing a limited amount of adenine (40% of the normal amount; Figure 3.5).

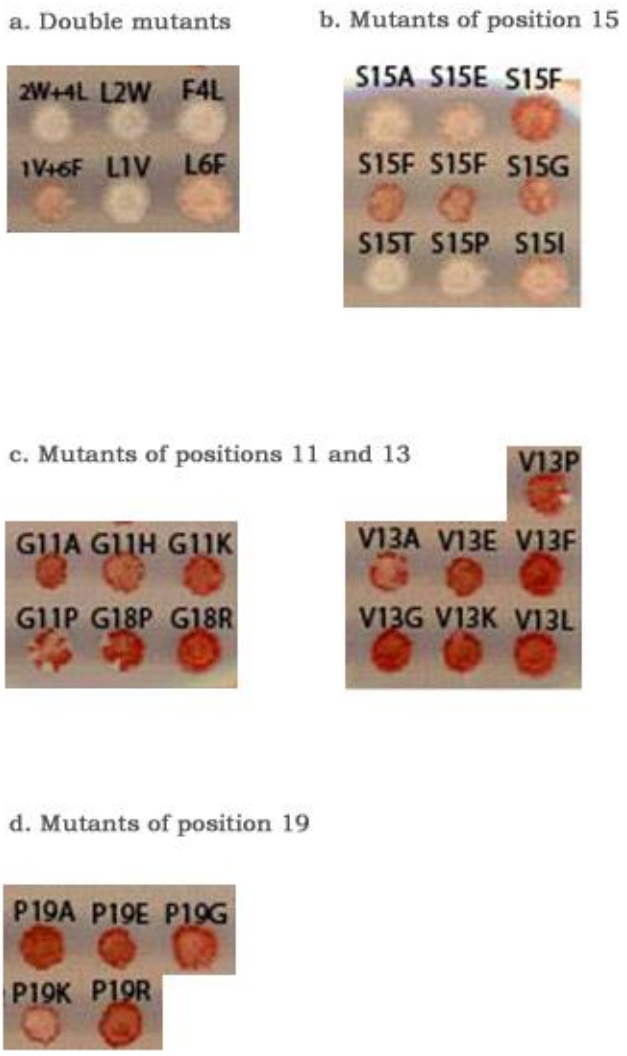


Figure 3.5 *in vivo* assay for the mutants listed in section 3.3. See text for detail. (Mutants G18R and G18P are not relative to the experiment, only occasionally appears in the certain experiment represent in the figure)

Examination of the double mutants (L1V+L6F and L2W+F4L) confirmed the previous observation that they have high activity, with L2W+F4L having similar activity to wild type. Comparison with 2A variants containing the single amino acid changes that comprise the double mutants (L2W, F4L and L1V, L6F) revealed that these also had

high levels of activity. Specifically, both of L2W and F4L appeared to have similar activity to the combined mutant (L2W+F4L). Therefore whether the high activity of L2W+F4L is due to neutral effects or co-operative effect of the 2 single changes need further assay to address (see below). On the other hand, L1V+L6F was less functional than L1V, and therefore the high activity this double mutant presented was largely due to the neutral effects of the 2 single changes rather than any co-operative effect.

The sequences of known 2A peptides suggest that the position 13 and 15 are not completely conserved. While 13th position is found only with valine or isoleucine, the 15th is more variable with quite a few different amino acids found naturally. This difference in conservation was reflected in the present mutagenesis study. Replacement of Val13 with any of the amino acids used here, including other hydrophobic amino acids (Leucine and phenylalanine) resulted in significant loss of 2A activity. In contrast, the effects of substituting Ser15 varied depending on the amino acid inserted. There was though clearly more flexibility at position 15 as none of the substitutions completely inactivated 2A, while others (such as serine, proline or alanine) had little, if any effect. This contrasted with the Tyr15 mutant isolated in the random screen which abrogates 2A activity.

While the majority of alterations at position 19 inactivated 2A, strikingly, mutant P19K retained some 2A function. Another mutant, P19A, though not always, also presented 2A activity in the majority of the experiments (however not in the experiment shown in figure 3.5). These results were significant in that previously Proline at position 19 had been considered to be necessary for 2A activity. Because the pigmentation of the colonies expressing these mutants was hard to quantify, being only slightly less dark than those with other mutations to P19, a further study using *in vitro* assay was needed to quantify it exactly (see below).

The substitution of Gly11 for any of the amino acids tested caused dramatic reduction of 2A function. This included histidine which is the second most common amino acid that is found in this position in known 2A peptides. Position 11 is the first conserved

position in 2A, and it may provide a ‘link’ between the N-terminal non-conserved, and C-terminal conserved portions of the peptide, a more specific analysis on this position was carried out below (see Chapter 5).

In general, the *in vivo* test described in this section was a good complimentary study to Dr. Sharma’s random screen, providing several interesting start points for further research.

3.3.2 in vitro test

As mentioned at last section, quantitative and more precise analysis was required for study of several interesting mutants. Therefore, the *in vitro* assay was performed with a set of mutants that included 2 double mutants (L2W+F4L, L1V+L6F) and their separated individual mutants, as well as P19K, P19A.

The total eight resulting mutants were inserted into the plasmid pLH135 and experiments were performed as described in section 3.2.3 and 3.2.2. Representative experiments and the quantitative result were presented in figure 3.6 and table 3.3

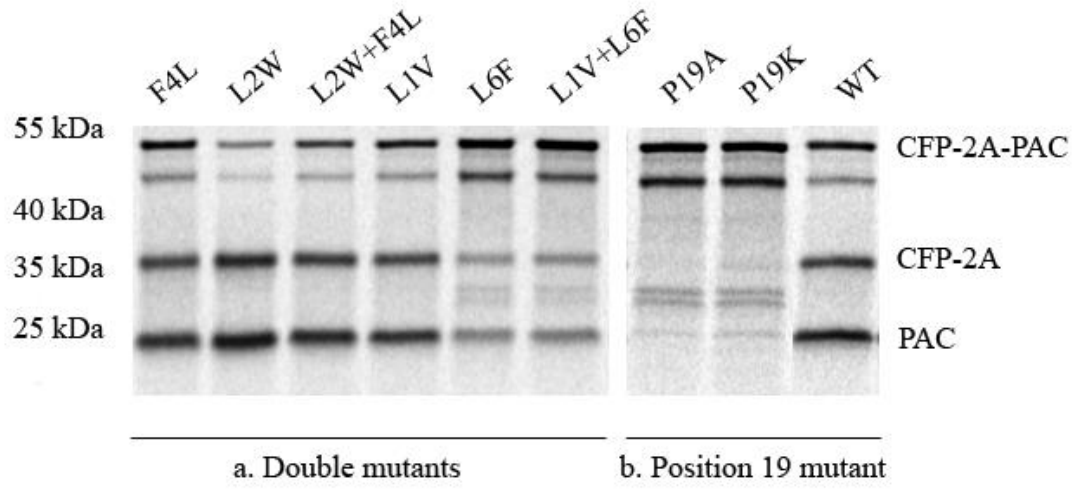


Figure 3.6: *in vitro* assay for the mutants listed in section 3.3.2 see text for detail

Mutants	Activity by *in vitro* assay	Colony colour *in vivo* assay
Wild type	70.8%+/-4.2%	White
F4L	61.2%+/-1.1%	White
L2W	94.4%+/-0.9%	White
F4L+L2W	79.5%+/-0.6%	White
L1V	67.2%+/-2.0%	White
L6F	22.7%+/-5.9%	Pink (light)
L1V+L6F	22.5%+/-6.5%	Pink (light)
P19A	4.5%+/-1.8%	Red (variable)
P19K	6.1%+/-1.1%	Pink (deep)

Table 3.3 Quantification of the *in vitro* assay. The experiments were performed and quantified as described in section 3.2.3 and 3.2.2. For one mutant three independent assays were applied, the results were summarized in column 2 'Activity by *in vitro* assay', in 'average value +/- standard deviation' form. The result of in vivo assay of these mutants obtained in last section were also listed for comparing, in column 3 'Colony colour *in vivo* assay'

Results obtained in the *in vitro* experiments were consistent with those generated with the *in vivo* reporter. The mutants conferring white colonies in *in vivo* test yielded high activity in *in vitro* analysis, whereas those that grew as red/pink colonies had low efficiency in the *in vitro* reporter.

In this *in vitro* assay, the double mutant L2W+F4L and the two single mutants that comprise it presented high activities. L2W gave 94.4%+/-0.9% efficiency, which was higher than that of wild type FMDV 2A (70.8%+/-4.2%), and was the highest 2A activity in whole study described in this thesis. The strong activity of L2W+F4L (79.5%+/-0.6%) was then not due to two mutant that individually reduce 2A activity complementing each other, but more likely attributed to the positive effect of the L2W mutant somehwat compensating for the relatively low activity of F4L with a relatively lower activity at 61.2%+/-1.1%.

For double mutant L1V+L6F, the *in vitro* assay confirmed the conclusion of *in vivo* test: the quantification data clearly showed L1V (67.2%+/-2.0%) single mutant having same level of activity as wild type 2A, the relative high level of the double mutant L1V+L6F (22.5%+/-6.5%) should mainly attribute to the fact that replacement of Leu to Val at the first position has little effect on its activity.

Two single mutations at Pro19 found to have a low level of activity in the *in vivo* reporter were also active in the *in vitro* experiments (Figure 3.5b showed the faint but clear bands of the up-and down-stream products) Quantification yielded 7% and 6% activity for P19A and P19K respectively. These results were consistent with previous theory of the importance of Pro19, though these data indicate that it is possible for the 2A reaction to proceed, albeit at low efficiency, with amino acids other than Proline at this position.

3.4 Summary and discussion

Site-specific mutagenesis study presented in this chapter is a good complimentary study of the previous random mutagenesis study. The analysis of Val13 and Gly11 filled the 'holes' of the random screen. The result of the analysis, combined with Dr Pamila Sharma's results, suggest the whole 19 amino acid in 2A sequence is more intolerant to mutation than expectation. Alteration at any of the 19 positions changes (mostly impairs) the 2A function. Thus, 2A is unlike some other short peptides, the whole 2A sequence is important and every amino acid contributes the function in some way.

Substitutions of glycine to histidine at position 11 caused dramatic loss of 2A activity. This result is slightly unexpected, as histidine is the second most frequent amino acid that is found at this position in identified 2A peptides. Since histidine at position 11 functions well in other 2A sequences, such as the one from TMEV (Donnelly *et al*., 2001), given the high conservation of the C-terminal motif, the element(s) that determine the functional amino acid at position 11 is mostly likely contained within the variable N-terminal portion, i.e position 1-10. The potential relationship between position 11 and N-terminal portion will be discussed in more details in chapter 5.

Valine and Isoleucine are the only two amino acids that are found at position 13. As this two amino acids share the properties of strong hydropathy and the branched side chain, these features were thought as the requirements of the amino acids at this position. However, present study revealed that, the substitution of the hydrophobic and/or branched side chain comprised amino acids (leucine, alanine, phenylanine and glutamic acid) displaced no different activity from these with other amino acids (proline, glycine, lysine): none of the mutation at position 13 tested in this study resulted in a significant activity. Although the mutations tested here is not an exhaustive set, this result suggests a more strict requirement at this position for a functional sequence.

In contrast to position 13, the position 15 of 2A sequence is more variable in known 2A peptides. Intriguingly, it was observed in previous random screen that, tryptophan

replacement at this position nearly inactivated the function of the peptide. (table 3.1 S15I) Therefore, the question arose as whether a broad or limited range of amino acids at this position is compatible with full 2A function, and to answer this question, Ser15 was further mutated with the replacements of A,E,F,G,T,P and I at this position. The result, though varied depending on the amino acid inserted, in general suggested a more flexible requirement at this position, as none of the mutants tested, except S15F that resulted in a medium level of function, induced a substantial impairment of the activity. The reason why tryptophan not allowed at position 15 requires further analysis to reveal.

Pro19 at the end of 2A is considered indispensable for 2A function. However, the present study revealed that mutant 2A sequence with other amino acids in this position still function in at a minor level. This result is consistent with an old study showing that TMEV 2A tolerates some substitution at Pro19 (Hahn *et al*., 1997). This implies that the mechanism of 2A is more complex than thought.

In previous study, two double mutants were noted because that their activities are more than one of the single mutants alone. Present project generated all of their single mutants and tested them with more precise assay. It was revealed that the strong activity of double mutants is more likely contributed by their other single mutants along, rather than the interaction within 2A sequence or compensation effects of two mutants.

In summary, the present site-specific mutagenesis study answers the several questions left by previous random screen study, and confirms the main finding of previous study: that the 2A sequence is NOT as flexible as one might imagine, particularly given the variability with in the N-terminus of the peptide.

Chapter 4: Systematic substitution of each FMDV 2A position

This chapter describes a systematic scanning mutagenesis of 2A. Each amino acid of FMDV 2A was individually replaced by alanine, proline or glycine and the activity of the mutants were assessed by *in vitro* and *in vivo* (partly) assay. There were two aims/reasons for this: the first is that it was a complementary approach to the previous random screen. The random screen combined with site-specific mutagenesis analysis described in chapter 3 had revealed that the mutations at every position of the sequence affected its 2A efficiency. However, while comparing the tolerance for mutation of different positions, the various properties of different amino acids substituting at different positions might affect the conclusion. Thus, analysis of a set of 2A mutants in which each position was substituted by a same amino acid could provide a new angle to examine the general trend of the effect of altering each position within the peptide. The second aim was to test the importance of secondary structure in 2A function. In previous studies, it was observed that more than half of 2A sequences identified, including the one from FMDV, have a strong α-helical propensity in their N-terminal portion, or indeed much of the whole sequence (Ryan *et al*., 1999). Since three amino acids selected for the scanning analysis, alanine, glycine and proline have distinct influences on secondary structure, comparing the difference in effects of these amino acids was helpful to assess the significance of α-helical propensity.

4.1 Analysis of 2A sequences: secondary structure propensities

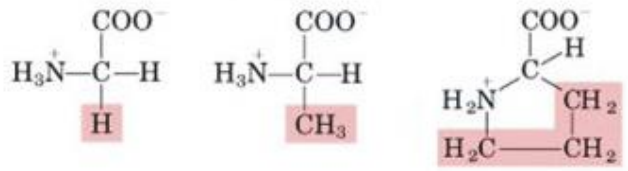
More than half of 2A sequences identified, especially these from picornavirus, have been suggested to adopt an α-helical conformation within the ribosomal exit tunnel over much of their sequence. In general, the most likely conformation of an amino acids chain inside the exit tunnel is α-helical (Nissen *et al*., 2000). All of the amino acids of the N-terminal part of these 2A peptides are favourable to form an α-helical structure; and on an earlier study (Ryan *et al*., 1994) using molecular dynamics modeling it was found that, the whole sequence of FMDV 2A except two final amino acids may adopt an α-helical conformation with a reverse turn at the C-terminal end.

As mentioned in the introduction, the 2A reaction is a co-translational event. When the event occurs, the two final codons of 2A, proline and glycine are in the A and P site of ribosome respectively, with the rest of the 2A peptide within the ribosomal exit tunnel. Abundant evidence suggests that newly-synthesized polypeptides are able to form secondary structure or even ternary structure (Kosolapov *et al*., 2009) within the tunnel. Indeed, forming a secondary structure inside the exit tunnel is not an unusual strategy for the peptides manipulating the translation procedure. During the SecM- or AAP-induced translation stall, the peptide inside exit tunnel is in a compact, likely helical form (See section 1.2.4.1) (Woolhead *et al*., 2006; Bhushan *et al*., 2010). Therefore, it is possible that during the 2A reaction, the growing amino acids chain of 2A adopts an α-helix conformation inside ribosome, with two final codons at the peptidyl tranferase centre. One of the aims of the study described in this chapter is to examine the roles and importance of this potential structure to the 2A event. For this purpose, three amino acids with distinct impact on α-helical conformation were selected for the systematic scanning, namely proline, glycine and alanine.

Glycine is one of most unfavorable amino acids to form the α-helix and proline is unable to adopt the correct conformation to ‘fit’ into an α-helix; indeed examination of known structures reveals that they have two lowest frequency of occurrence in α-helical

conformation (Figure 4.1b). These two amino acids obtain their ‘helix breaker’ feature by distinct mechanisms: proline has a unique, cyclic side chain (proline ring) that connects with the peptide backbone twice (Figure 4.1a). This constrains the confomrations that the backbone can take and prevents the formation of α-helix. In contrast a peptide backbone containing glycine, without a side chain, is easily reduced to coil. Alanine, with a non-polar ($–CH_3$) side chain has a high frequency of occurence in α-helices and insertion of alanine in place of other amino acids typically does not disturb an existing helix (Figure 4.1b). Due to these different properties, it might be expected that the effects of the substitution of individual amino acids of 2A with proline, glycine and alanine may be different, and the significance of α-helical propensity can be addressed by this variability.

Figure 4.1a Three amino acid picked for scanning study



Glycine Alanine Proline

Figure 4.1b amino acids α-helical propensity scales

AA	Exptl	AK/AQ	AGADIR	Williams	C. & F.	Luque
Ala	0	0	0	0	0	0
Arg	0.21	0.22	0.06	0.20	0.44	0.15
Leu	0.21	0.37	0.19	0.07	0.21	0.15
Met	0.24	0.52	0.21	0.11	−0.03	0.18
Lys	0.26	0.29	0.15	0.18	0.26	0.11
Gln	0.39	0.55	0.32	0.14	0.31	0.30
Glu	0.40	0.62	0.34	−0.18	−0.09	0.37
Ile	0.41	0.71	0.35	0.32	0.34	0.48
Trp	0.49	0.96	0.47	0.39	0.34	0.35
Ser	0.50	0.79	0.52	0.84	0.65	0.48
Tyr	0.53	0.69	0.47	0.67	0.73	0.46
Phe	0.54	1.00	0.47	0.25	0.29	0.35
Val	0.61	1.04	0.46	0.43	0.36	0.36
His	0.61	0.97	0.62	0.36	0.42	0.62
Asn	0.65	0.96	0.6	0.65	0.75	0.52
Thr	0.66	1.22	0.57	0.65	0.59	0.59
Cys	0.68	0.91	0.6	0.75	0.72	0.57
Asp	0.69	0.81	0.59	0.42	0.41	0.47
Gly	1.00	1.97	1.10	0.98	0.85	0.79

Figure 4.1 distinct α-helical propensities of glycine, proline and alanine.

(a) Chemical structure formula of the three amino acids. The different side chains should be noted. (Taken from "Lehninger principles of biochemistry")
(b) α-helical propensity scales of amino acid, by different methods. Exptl, AK/AQ (Rohl *et al*., 1996), based on the data from helix-to-coil equilibrium experiments; AGADIR (Munoz *et al*., 1995), by the data from peptides; Williams (Williams *et al*., 1987), C&F (Chou *et al*., 1978), based on the frequency of occurrence of amino acids in a-helices in proteins; luque (Luque *et al*., 1996) based on the structure biology study. The data presented as D(D*G*) values relative to Alanine, which has been set to zero because it is usually the amino acid with the most favourable helix propensity. Proline is not in the scale due to its extreme non-preference to the α-helical structure (reproduced from (Pace *et al*., 1998).

4.2 Scanning mutagenesis of FMDV 2A and *in vitro* analysis

To accurately assess and compare the 2A function of different mutants, the 2A activities of the variants in this scanning study were mainly assessed with the [CFP-2A/2A*-PAC] reporter and *in vitro* test. 47 target mutants, comprising 18 mutants in which each amino acid of FMDV 2A was individually replaced by alanine, 17 that of proline and 13 glycine mutants, were generated by the site-specific PCR using the primers specifying the desired change, as described in section 3.2.3.

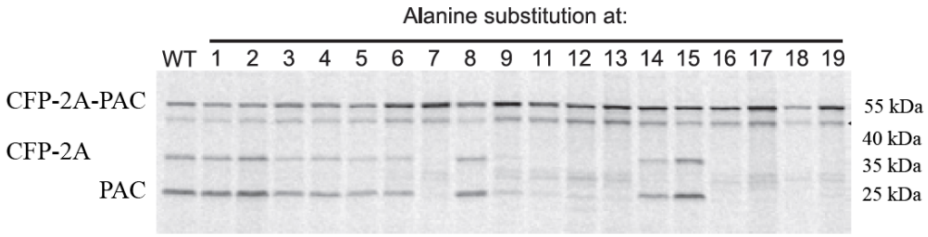
The other four glycine mutants, namely D5G, L6G, A10G, E14G, were obtained by different approach. These mutants were isolated in Dr Pamila Sharma's random mutagenesis study, (section 3.1) therefore they could be obtained from the collection of PCR fragments used to identify the mutant in previous study. For exploiting these fragments, a pair of primers (Methods and Materials Ubiq5 and AD3 in table 2.3) that complemented the flanks of these fragments was used to amplify them from the existing

constructs (Method and Materials section 2.4.3).

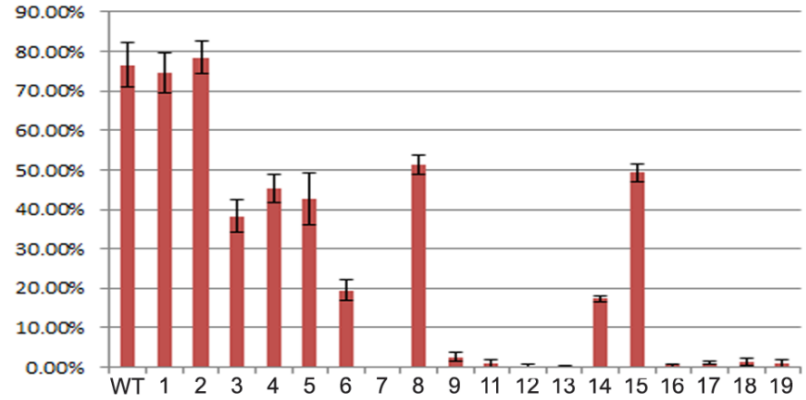
The 52 PCR products corresponding to the unique 2A mutants were then inserted into plasmid pLH135 plasmid using XbaI and ApaI restriction sites, replacing the wild type 2A coding sequence of the [CFP-2A-PAC] reporter with the mutant ones. The resulting plasmids were sequenced by GATC biotech with primer 135SEQ, to conform the present of desired mutations, as well as no undesired mutations introduced during the procedures.

The sets of mutants were then used in coupled *in vitro* transcription/translation assays, as described in section 3.1.2. Representative gels for the glycine, proline and alanine series of mutants (a) and quantification data of the assay (b,c) result are shown in Figure 4.2-4.4.

a. Representative experiment



b. Quanfication Data



c. Quantification Data (set FMDV wild type as 100%)

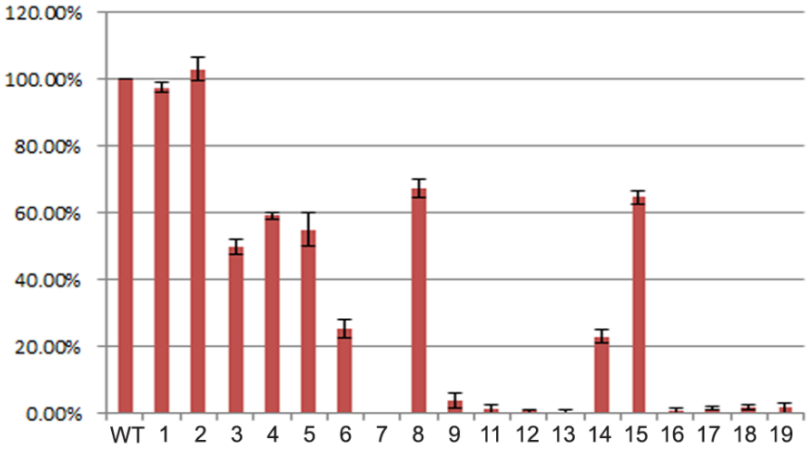


Figure 4.2 Activities of [CFP-2A-PAC] variants containing Alanine substituted for the individual amino acids of 2A indicated, and the wild-type (WT). See text for details.

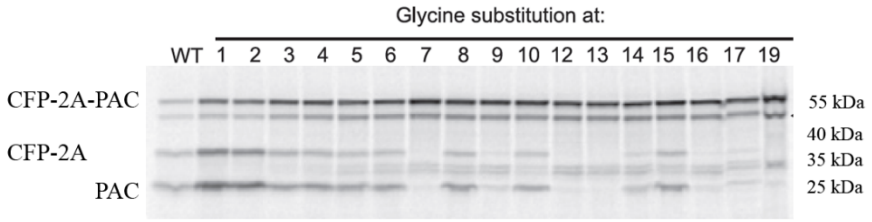
a. A representive *in vitro* assay of the whole set of alanine substituted 2A mutants. E, up and down refer to the complete (extended) translation product [CFP-2A-PAC], and the

upstream [CFP-2A] and downstream [PAC] products of the 2A reaction, respectively.

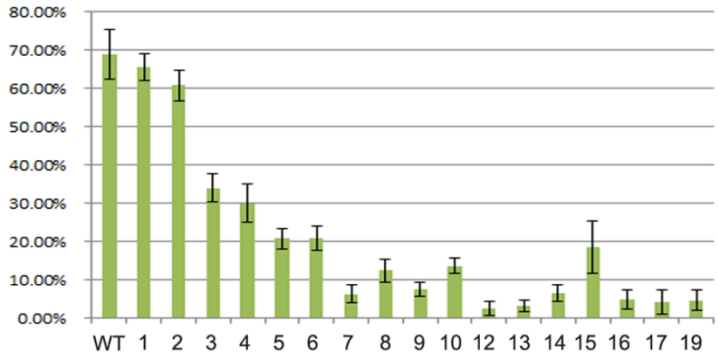
b. Three independent sets of experiments as in a. were quantified for each set of substitutions. Average activity (% of ribosomes releasing the upstream product) ± one standard deviation is shown.

c. Same as that in b., but 2A activity was seted as their ratio to the acitvity of wild type 2A.

a. Representative experiment



b. Quanfication Data



c. Quantification Data (set FMDV wild type as 100%)

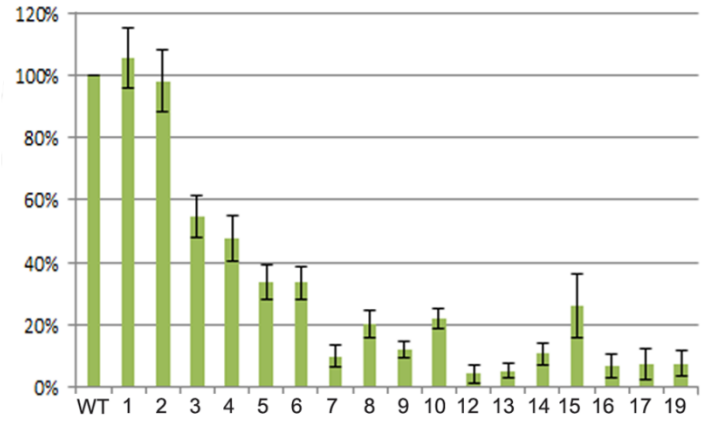
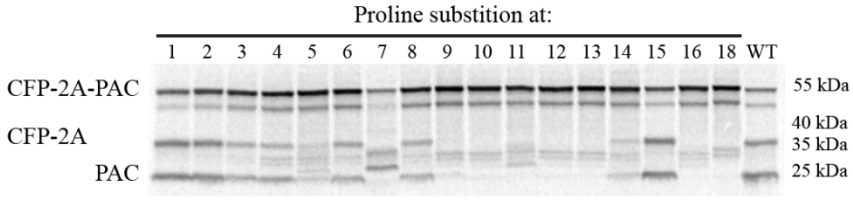
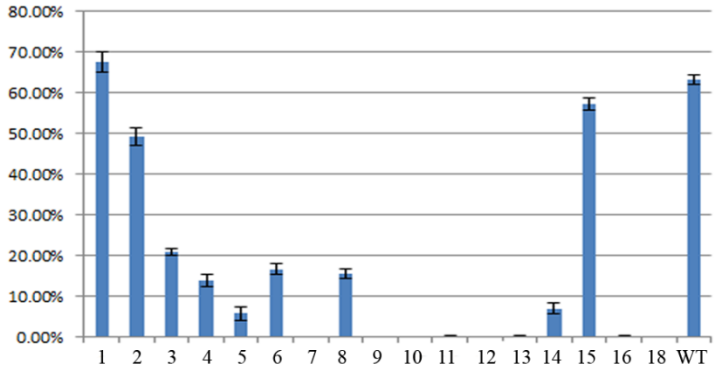


Figure 4.3 Activities of [CFP-2A-PAC] variants containing Glycine substituted for the individual amino acids of 2A indicated, and the wild-type (WT). See text for details. Figure 4.3 a,b,c are in the same form as Figure 4.2.

a. Representative experiment



b. Quanfication Data



c. Quantification Data (set FMDV wild type as 100%)

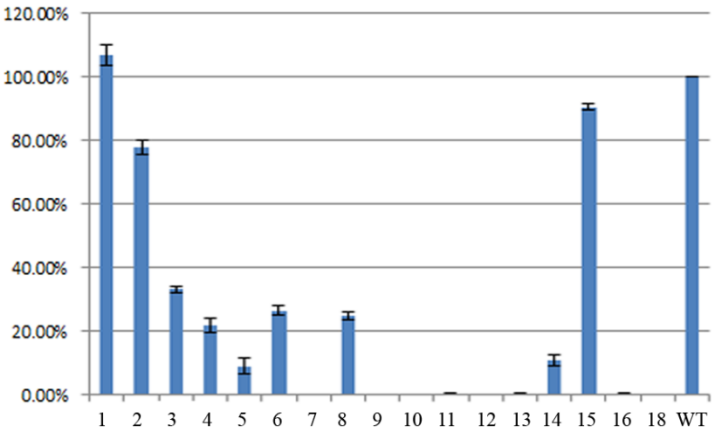


Figure 4.4 Activities of [CFP-2A-PAC] variants containing Proline substituted for the individual amino acids of 2A indicated, and the wild-type (WT). See text for details. Figure 4.4 a,b,c are in the same form as Figure 4.2

Individual replacement of each amino acid of the 2A peptide with glycine, proline or alanine led, in general, to impaired activity. Out of a total of 52 unique mutants tested in this scanning study, only three substitutions, L1P, L1G and L2A, that were neutral; and four, L2P, L1A, L2G, and S15P which resultsed in only minor loss of the activity. The rest 45 mutants suffered significant loss of the activity, and 6 of 7 of these exceptions occurred in Leu1 and Leu2. Hence, this result was consistent with the conclusion of previous studies based on random screen, that whole sequence of 2A peptide is largely intolerant to mutagenesis.

The effects of altering each position to glycine, proline, and alanine are generally similar. In most cases, substitutions to amino acids more C-terminal within 2A had more deleterious effect than changes to those within the N-terminal part, with a main exception of position 15. While general trends were consistent with the previous alignment study suggesting the conserved C-terminal and variable N-terminal portion (Donnelly *et al*., 1997), this result was unexpected to the distinct properties of glycine, proline, and alanine, as mentioned above. Given these differences of the three amino acids, especially in term of how they influence the conformation of the nascent chain, the data obtained argue against specific secondary structure of a large part of 2A (especially the N-terminal portion) being necessary and sufficient for the 2A reaction. In particular, since substitutions to alanine had similar effects to glycine and proline the data argue against an α-helical conformation being sufficient. Therefore, the similarity of effects of their substitutions could only be explained that, the α-helical propensity or potential α-helix structure alone could not satisfy the requirement(s) for 2A function, there must be other factors such as side chain involved.

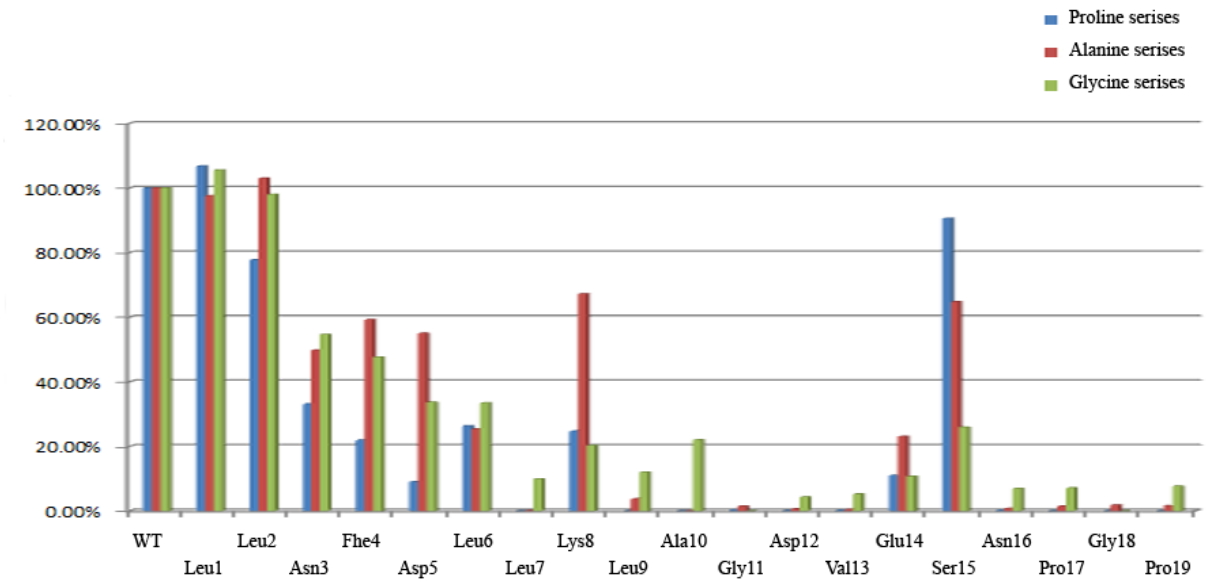


Figure 4.5 Comparison of three amino acid scanning. At each position, the 2A activity (ratio to the wild type) of mutants containing proline, alanine and glycine are summarized from Figure 4.2c, 4.3c and 4.4c.

Although the general patterns of Glycine, Proline and Alanine were similar, indeed they gave some minor differences: on the whole, at some position, substitution with Proline cause a more several reduction of 2A activity than the other two, with the exception of position 15. Moreover, several differences were observed at some certain positions. At positions 4-5, alanine substitutions showed relatively considerable 2A activities (F4A 55% and D5G 59% as wild type). This had a minor edge to the glycine series (F4G:42% D5G:28%) and a pronounced advantage to proline series (F4P:21% D5P:9%) at the same positions. Therefore, while the α-helical propensity might not be that critical for the most of the length of 2A peptide, it might be required in this region of the sequence. However, the amino acids at these positions in wild type FMDV, phenylalanine and aspartic acid, are the two of only three amino acids within the N-terminal portion that are relatively unfavourable to the helix. This fact discounts the above argument. A more

promising difference was shown at position 8, where the alanine mutant displaced pronounced to both glycine and proline ones. Another position that showed obviously different effect was position 15, the only position that was highly variable in the C-terminal motif. At this position 15, it was the glycine replacement S15G that resulted in the lowest function (only 25%). The activity of S15P and S15A were 90% and 65% respectively. The substitution of glycine impaired the function more severely than that of others. This might be due to the detrimental flexibility of the peptide backbone introduced by glycine, or another possibility of the 15th position's requirement for side chain size of amino acid.

4.3 *In vivo* examination of scanning mutants of FMDV 2A

The *in vivo* analysis of scanning mutants of 2A was a supplementary study to the *in vitro* scanning described in section 4.2. The main aim of these experiments was to check whether the conclusions of the *in vitro* scanning were independent of the organism and the method used. In *in vivo* scanning, PCR fragments containing 26 different mutants, including the whole glycine series and 9 proline mutants (at position 1,2,4,7,10,11,13,15,18) were selected from the same collection of 52 PCR products used in *in vitro* study, and placed into the *in vivo* reporter [UBI-R2A-ADE2] plasmid pJNY58, by NcoI and BamHI sites. The resulting plasmids were sequenced with primer AD3 for conforming, and then transferred into yeast strain JDY4 for *in vivo* assay, as described in section 3.1. Results of a typical experiment is presented in Figure 4.6

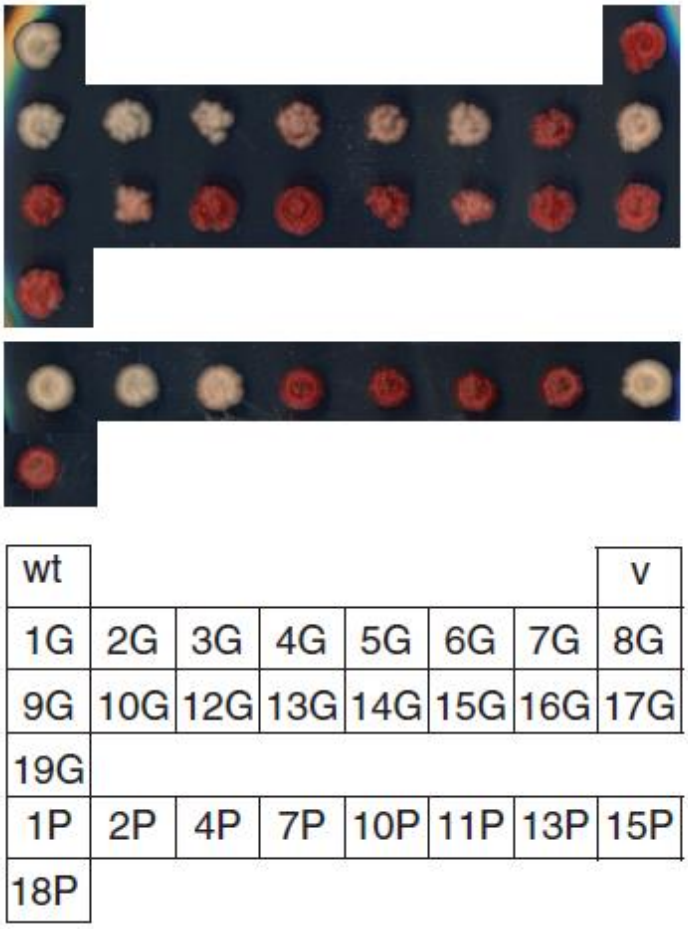


Figure 4.6 *in vivo* assay of whole glycine series and part of proline series. See text for details. My data, figure reproduced from Sharma *et al*., (2012).

All mutants tested in this experiment yielded comparable results to those obtained *in vitro*. The full set of glycine substitution presented the same general trend as the *in vitro* assay, as did the selected proline mutants. These results confirmed that the functions of the 2A mutants showed in last section were irrelevant to the constructed used (pLH135 or pJNY58), method tested (*in vitro* or *in vivo*) and organism based on (mammal or yeast).

This result was the other evident that the *in vivo* and *in vitro* assay we used in this study complemented to each other.

4.4 Summary and Discussion

In the research described in this chapter, systematic substitutions were applied to the 2A sequence from FMDV with three different amino acids glycine, proline and alanine. The full sets of mutants of glycine, proline and alanine series were examined with *in vitro* assay; and whole series of glycine as well as 7 proline mutants were also tested with *in vivo* reporter system.

With the *in vitro* test, all three amino acids tested presented generally similar effects at each position of 2A position. In general, the closer, a position is to the C-terminal end, the more serious impairment that results from the substitution, with a major exception of position 15, at which the substitutions with proline and alanine retain considerable activity, and substitution with glycine showed a low level activity, but higher than that of glycine substitutions at surrounding positions. These results, combined with the studies based on random screen, suggests that this trend reflects general sensitivity of the positions of 2A to the mutations, regardless of different amino acids substituting, which, in turn, implies the significance of the amino acids located at these positions: while every amino acids of the whole 2A sequence contributes to the function, the contribution of those near the C-terminal end is more critical than that of N-terminal for 2A function. This, once again, coincides with previous 2A alignment study (Donnelly *et al*., 1997).

As the three amino acids selected here have distinct effects on secondary structure, the

similarity observed from the mutants of glycine, proline and alanine was unexpected. Due to alanine substitutions (that are favorable to the α-helical conformation) failed to show significant 'edge' over that of helix breaker proline and glycine at most of the sequence, the α-helical propensity previously thought critical for the 2A function might not be crucial, or at least that the structure of the backbone alone could not satisfy the requirement(s) of 2A function at most positions, other factors such as side chain must be considered.

As discussed in introduction, the structures of other peptides driving the ribosome vary: for example, the TnaC adopt an extended conformation inside the ribosomal exit tunnel (Seidelt *et al*., 2009), and the arginine attenuator peptide is in a compact form. Hence, the potential conformation of 2A peptide adopted within the ribosomal exit tunnel and its contribution to the 2A function needs further research.

The 15th position of the 2A sequence is the only position that is highly variable in the C-terminal motif. In the present scanning study, position 15 gave the greatest variation in different amino acid substitutions: S15G showed far less 2A function than S15P or S15A. A possible explanation for this might be that the flexibility introduced into the peptide backbone of 2A by glycine is unfavorable in this position. But after review of all the mutants tested in Dr. Pamila Sharma's random screen (see section 3.2) and the study described in this thesis (see section 3.3), an alternative requirement emerges. Of ten mutants occurred at position 15, amino acids with a similar sized side chain (molecular weight ranging 30 g/mol to 50 g/mol) caused the least effect on 2A activity, and it seems like that the more MW of side chain deviated from this range, the more serious impairment of the 2A function the amino acid substitution suffered. Indeed, while mutants containing proline, cysteine or threonine at 15th position exhibited similar high activity to wild type, alanine at this position (small side chain) yielded compromised activity; and tyrosine, with the bulkiest side chain, decreased 2A function dramatically. Therefore, there might be a requirement of the size of side chain for the position 15 of FMDV 2A.

Finally, the *in vivo* test based on yeast genetic method was performed to the selected mutants of the *in vitro* scanning. The results of these two tests were consistent with each other, proving that these data were independent to the construct, organism and the method used.

Chapter 5: Combination mutants of different type of 2A sequences

The 11th position of 2A sequence is the first amino acid that is conserved. In all 2A peptides that have been identified, glycine is the most common amino acid at this position. 80.6% of 2A sequences, including the one from FMDV, have glycine at position 11. The second most common amino acid at position 11 is histidine, which 14.5% of 2A sequences contain. Interestingly, site-specific mutagenesis revealed that, a histidine replacement to the position 11 of the FMDV 2A caused nearly total loss of its activity (section 3.3). Moreover, when comparing different 2A peptide it was observed that, the 2A sequences with Gly11 and these with His11 had distinct preferences for the amino acid found within the N-terminal 'non-conserved' region. Therefore, according to the amino acid at 11th position, 2A peptides can be classified into two types: Gly11 type and His11 type. As two types of 2A sequence appear to have very different constraints within the N-terminal portion, there may be some amino acids or combination of amino acids among N-terminal portion that interact with the amino acid at position 11. For identifying these potential functional units of N-terminal portion, two approaches were applied in the study described in this chapter. Firstly, the N-terminal half and C-terminal half (including position 11) from different type of 2A sequences were fused, to check whether they could function with each other; secondly, several 'snippets' (amino acids or combinations of amino acids), were picked from the His11 type 2A sequence, then fused into the corresponding positions of the FMDV mutant G11H. If a small portion of the H11-type 2A could restore activity within the G11H mutant of FMDV, this would delimit the interacting parts of the peptide.

5.1 Different type of 2A sequence present distinct constraints at 'non-conserved' positions

To aid in interpretation of mutagenesis data, an extensive alignment study was performed with 2A peptides from different viruses. During the period of the present study, our laboratory examined the full set (at that time) of identified 2A peptides (52 individual ones) (Sharma *et al*., 2012). Since then, several more 2A sequences have been discovered. Thus, here an updated alignment analysis is presented, with the same approach as the published data but a new collection of 2A peptides that is showed in (Table 5.1).

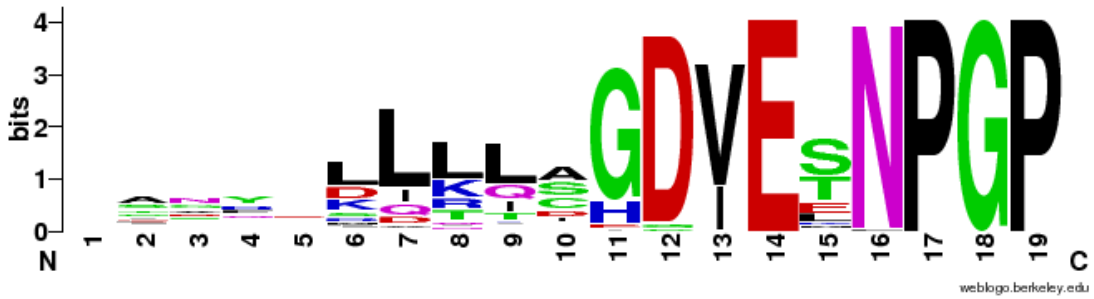
Organism	Sequence
Ljungan virus 145SL	YSGGKFLNQCGDVESNPGP
Porcine teschovirus	ATNFSLLRQAGDIEENPGP
Human cosavirus	QVARMLLQISGDVESNPGP
Theiler's encephalomyelitis virus	HADYYKQRLIHDVETNPGP
BHuV-1/OHuV-1	ATNFSLLRLAGDVELNPGP
SPaV-1	EELQNMILQCGDVEQNPGP
RaPV-1	GYRGDLLMQAGDVESNPGP
BRAV-2	VSNKDLLLQAGDVETNPGP
MsPV-1	MEDHSDILLGGDVEENPGP
Branchiostoma floridae	VISQLMLKLAGDVEENPGP

Table 5.1 The 2A peptides identified since August of 2011. To identify the new 2A peptides, the sequence information of virus updated within the year (form databases 'Picornaviurs home' (*www.picornaviridae.com*) and NCBI nucleotide collection/non-redundant protein sequences) was searched by the 2A C-terminal motif, then the resulting sequences were compared with the existing set of 2As to remove the

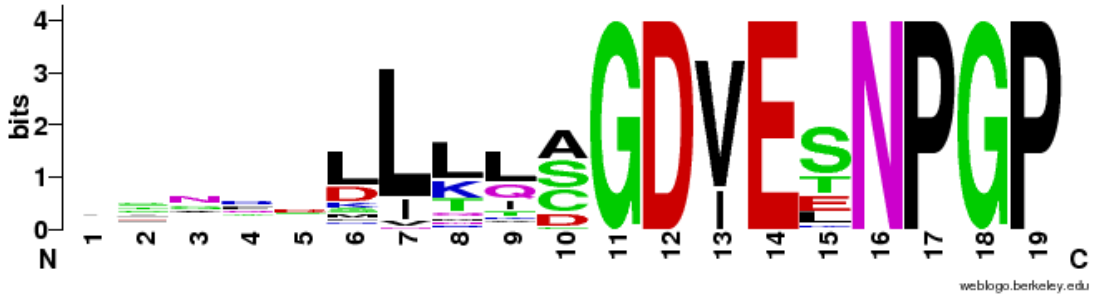
redundant ones. Finally ten putative new 2A sequences were identified and added into the existing collection of 2A peptides.

The updated 2A collection of total 62 (52 published in Sharma *et al*., (2012), plus 10 identified since 2011) unique 2A sequences as well as its two subsets, that of the sequences having glycine (of 50 sequences) or histidine (of 9 sequences) at position 11, were analyzed and represented by weblogos program (http://weblogo.berkeley.edu/logo.cgi), and the result of the analysis was displayed in Figure 5.1

A



B



C

Figure 5.1 Conservation within 2A peptides is shown as Weblogo representations. Each stack of letters corresponds to one position within the sequence,with the height of each stack indicating the conservation at that position and the height of letters within the stack the frequency at which the corresponding amino acid is found.

Separate Weblogos are shown for (A) the full set of 62 2A peptides (B) 2A peptides with Gly at position 11 (50 sequences) and (C) 2A peptides with His at position 11 (9 sequences).

As expected, the result of the new 2A collection was largely consistent with the original published one. Beside the C-terminal DV(I)EXNPGP motif, position 11 also displayed high conservation. Glycine was presented at this position in 50 of the 62 distinct sequences, and histidine in all but three of the rest, and the sequences with Gly11 and His11 displayed distinctly different amino acids in the N-terminal region as well as at position 15. In positions 6-9, leucine was the most common amino acid with Gly11, but comparably rare with His11, except position 9. Isoleucine is the most frequently found amino acid at 10th position of His11 type but is not seen in this position of Gly11 sequences. At the highly variable 15th position His11 2A peptides largely contained threonine or methionine, of which the latter never present in sequences that contain Gly11. These result, together with the finding that the substitution of glycine to histidine at position 11 of FMDV caused almost complete loss of activity (section 3.3), proposed the possibility there was co-operation(s) between amino acid at position 11 and these at the variable positions, and Gly11 and His11 held respective requirement over these amino acids.

Since these outcomes highlighted the significance of G11H mutant, here I tested this

mutant with *in vitro* assay to get the precise activity. The assay was performed with the method described in section 3.1, using the construct pLH135 built with the G11H containing PCR fragment. As expected from *in vivo* test, a very limited activity (9.1%) was found. (shown in Figure 5.3)

5.2 Fusion of two type of 2A sequence

To identify the non-conserved positions that co-operated with position 11, a set of 2A fusion sequences were made. Here, two sequences of His11 type of 2A, which were from Encephalomyocarditis Virus (EMCV) and Theiler's murine encephalomyelitis virus (TMEV), were chosen and combined with the sequence of FMDV 2A, which was of Gly11 type. These fusion mutants were placed into the pLH135 plasmid, and the result of *in vitro* assays using these is presented in this section.

5.2.1 EMCV and TMEV 2A sequences function with the in vitro reporter system

The 2A sequences of EMCV and TMEV presented high levels of 2A activity (102% and 77% when compared to FMDV 2A) in previous research using reporter construct [GFP-2A-GUS] and *in vitro* transcription/translation assay with wheat germ extract (Donnelly *et al.,* 1997). Here, we tested their function with the *in vitro* reporter system used in present study, the construct [CFP-2A/2A mutant-PAC] (pLH135) and coupled *in vitro* translation with rabbit reticulocyte lysate, to check whether they function with the system used in present project.

Due to the considerable variances between 2A sequences of EMCV, TMEV and FMDV, the strategy used to replace the wild type FMDV 2A with target sequence was different from that described in section 3.2 (see Figure 5.2). Here, for each sequence, two complementary oligonucleotides were designed that, after annealing, generated a fragment that coded for the desired 2A peptide, and also had 5' overhangs complementary to XbaI and ApaI sites at the 5' and 3' end of the fragments. These were then ligated into

pLH135 digested with XbaI and ApaI. The resulting plasmids were sequenced to confirm that they had the correct, and were then tested by coupled *in vitro* translation with rabbit reticulocyte lysate.

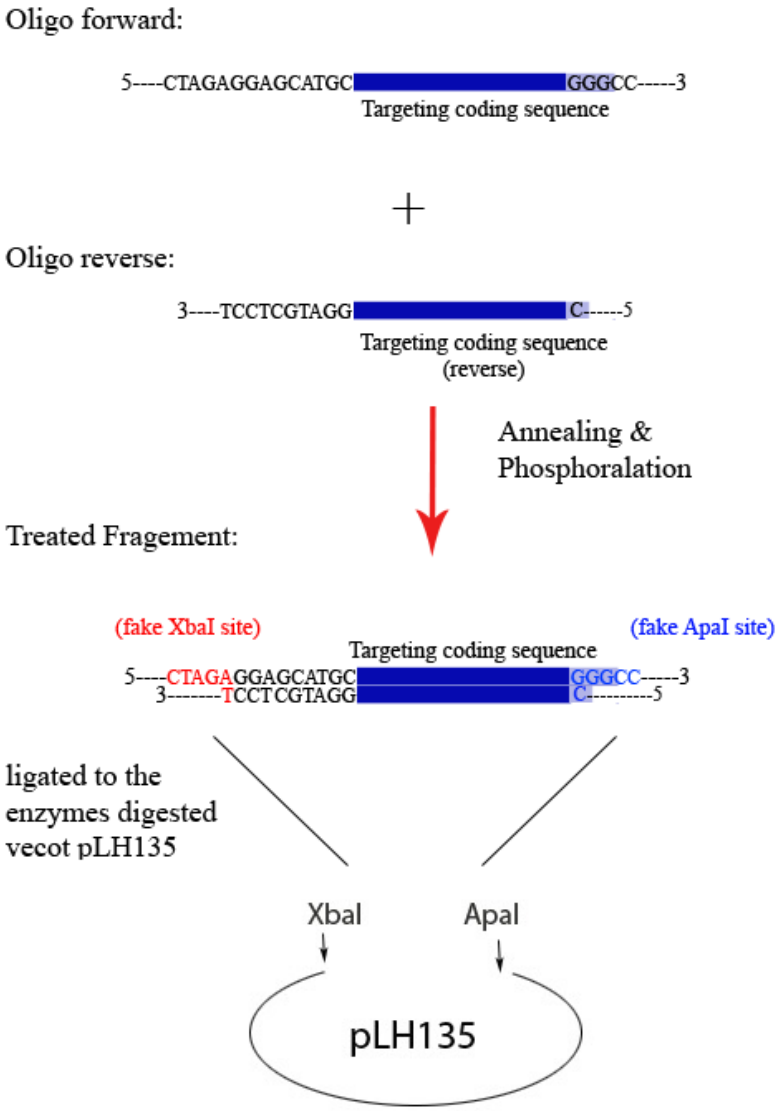


Figure 5.2 The clone method for the generation of mutants and assembling of pLH135 variants. Two oligoes containing the complimentary coding region that compose the target sequence and overhangs that compose two 'fake' restrict sites were designed.

After annealing as protocol in section 2.4.5, the resulting fragment was ligated to the enzymes digested vector pLH135.

The result of *in vitro* assay for EMCV and TMEV is shown in Figure 5.3. Quantification of this and similar gels is presented in Table 5.2 (Activity-present).

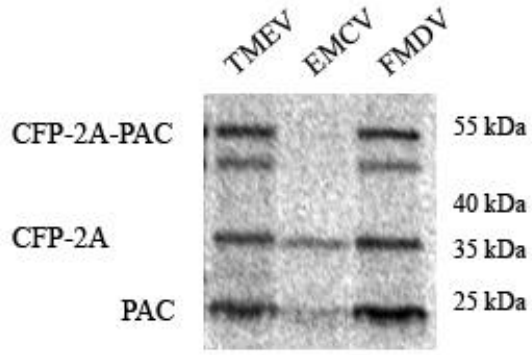


Figure 5.3 Representive experiment of *in vitro* assay for TMEV, TMCV and FMDV. Plasmids encoding [CFP-2A-PAC] reporter containing the indicated 2A sequence were used in coupled *in vitro* transcription/tranlsation assays, and resolved on a SDS-PAGE gel. Migration of markers run alongside are indicated.The assays were performed with the same protocol described in section 3.1.2, chapter 3.

2A species	Activity-present	Activity-previous
TMEV	60.66% +/- 6.80%	65%
FMDV	70.04% +/- 0.74%	90%
EMCV	92.84% +/- 1.87%	91%

Table 5.2 Quantified activities of *in vitro* assay for TMEV, TMCV and FMDV. For each 2A sequence, the average activity and standard deviation of 3 experiments was listed in column 'Activity-present'; and previous quantified data obeserved in (Donnelly *et al*., 1997) were listed in 'Activity-previous' for comparison.

The 2A from EMCV and TMEV were active, in the system of present study, and particularly, the quantification data obtained here were comparable to the data of the previous study. (Donnelly *et al*., 1997) Due to different methods and quantification formulas used the specific values of the two studies were not identical, but the relative activities were consistent to each other. (Table 5.2) It was concluded that 2A peptides form EMCV and TMEV were functional in the *in vitro* system used in present project, and were then a good start point for the study of the requirements for H11- and G11-type 2A peptides to function.

5.2.2 Fusion mutants of EMCV, TMEV and FMDV of 2A sequences are largely inactive.

To examine potenital link(s) between the N-terminal portion and C-terminal motif of 2A,

and particularly the amino acid at position 11, five fusion mutants of the N-terminal half and C-terminal half from different 2As were made. The fusion mutant was named after the first letter and the number of amino acids of two 2A sequences fusing it. For example, '10T+9F' mean a mutant fused by the first 10 amino acids of TMEV 2A peptide and last 9 of that of FMDV. The names and sequences of the five fusion mutants analysed in present study are shown in table 5.2. The approach described in section 5.2.1 was used to generate these mutants.

Coupled *in vitro* transcription/translation assays with the constructs bearing the 2A fusions revealed (Figure 5.4) that they had limited 2A activities (from 13 to 25% of FMDV 2A). Since each of the native 2A peptides functioned well,-the initial conclusion was that the impairments of activity must be ascribed to failed co-operation(s) between the N- and C-terminal halves. Close examination of the differences between native 2A peptides and the fusions generated suggests that their may be interaction between position 11 and 15. Specifically, the fusion T10/11F is equivalent to a H11G/T15S double mutant of the TMEV 2A. However, the function of a 2A peptide with histidine at position 11 is not entirely due to the presence of threonine (or methionine) at position 15, otherwise the fusions of the FMDV N-terminal portion to the C-terminal regions of TMEV or EMCV 2A would have been expected to function well, which they did not. This this, indicates the existence of interaction(s) between glycine/histidine at position 11 and the rest of 2A.

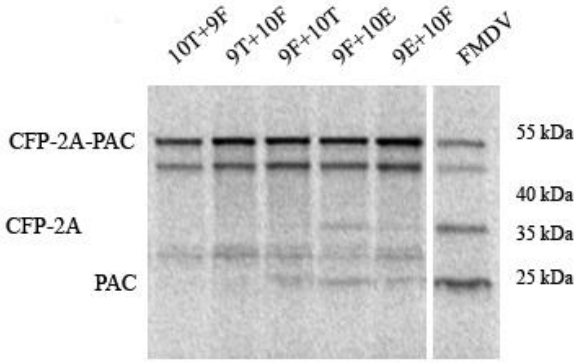


Figure 5.4. Representive experiment of *in vitro* assay for fused 2A mutants. Plasmids encoding [CFP-2A-PAC] reporter containing the indicated 2A sequence were used in coupled *in vitro* transcription/tranlsation assays, and resolved on a SDS-PAGE gel. Migration of markers run alongside are indicated. Note that the FMDV sample was run on the same gel but not alongside the other samples. The assays were performed with the same protocol described in section 3.2.2, chapter 3.

Mutants	Sequence	Activity
TMEV	HADYYKQRLIHDVETNPGP	60.66%+/-6.83%
FMDV	LLNFDLLKLAGDVESNPGP	70.04%+/-0.74%
EMCV	YAGYFADLLIHDIETNPGP	92.84% +/- 1.87%
10T+9F	HADYYKQRLIGDVESNPGP	11.90%+/-3.27%
9T+10F	HADYYKQRLAGDVESNPGP	9.84% +/-2.68%
9F+10T	LLNFDLLKLIHDVETNPGP	14.48%+/-1.11%
9F+10E	LLNFDLLKLIHDIETNPGP	18.16%+/-4.43%
9E+10F	YAGYFADLLAGDVESNPGP	15.00%+/-2.88%

Table 5.3 Sequence and quantified activities of *in vitro* assay for fused 2A mutants. For each 2A mutant, the average activity and standard deviation of 3 experiments are listed in column 'Activity'; the sequences of the fused mutants are presented in column 'Sequence', the sequence of TMEV, FMDV, and EMCV is highlighted with colour blue, black, and red respectively.

5.2.3 Inserting amino acids from His11 type 2A peptides into the G11H FMDV mutant failed to recover 2A function

The results described in the last section suggested that co-operation and communication within amino acids of 2A, especially between the glycine/histidine at position 11 and the amino acids located in the N-terminal portion of the peptide. However, the particular amino acids or combination of amino acids that responded to the position 11 were yet identified. For pinpointing this, a variety of changes were made to the G11H FMDV 2A mutant. Mutants made are listed in Table 5.4, and reflect the most frequently occurring amino acids in H11 type 2A peptides. Mutations were also inserted into the wild type (G11) FMDV peptide to test whether they interfered with function. This was an important control given that most changes to FMDV 2A affect its function (Chapters 3 and 4).

Positions 4-7 are most commonly YYKQ in H11 type 2A peptides and it was noted that these amino acids are rarely found in the same positions in G11 type 2A peptides, and thus this stretch of amino acids was considered to be a good candidate for allowing a H11 peptide to function. Following the same logic (common with His11 but rare with Gly11) as above, A10I, and L7D mutants were also examined in the experiments described in this section.

The mutants mentioned above were generated as section 3.1, except for YYKQ where the method described in section 5.2 was used. After constructing and sequencing, these mutants were tested by *in vitro* assay, of which the quantification data were presented in

table 5.4.

Mutants	Sequence	Activity
G11H	LLNFDLLKLAHDVESNPGP	9.1%+/- 2.7%
A10I	LLNFDLLKLIGDVESNPGP	5.5%+/-7.3%
A10I+G11H	LLNFDLLKLIHDVESNPGP	6.2%+/- 2.8%
L6K	LLNFDKLKLAGDVESNPGP	33.2%+/- 2.7%
L6K+G11H	LLNFDKLKLAHDVESNPGP	1.3%+/- 2.2%
L7Q	LLNFDLQKLAGDVESNPGP	2.5%+/- 1.8%
L7Q+G11H	LLNFDLQKLAHDVESNPGP	3.1%+/- 0.9%
L7D	LLNFDLDKLAGDVESNPG	0.9%+/-1.5%
L7D+G11H	LLNFDLDKLAHDVESNPGP	3.1%+/-3.1%
KQ+G11H	LLNFDKQKLAHDVESNPGP	1.2%+/- 2.4%
YYKQ	LLNYYKQKLAHDVESNPGP	0.2%+/- 0.3%

Table 5.4 Sequence and quantified activities of *in vitro* assay for 2A mutants mentioned in section 5.2.3. For each 2A mutant, the average activity and standard deviation of 3 experiments are listed in column 'Activity'; the sequences of the fused mutants are presented in column 'Sequence', the sequence of FMDV and the sequence fragment common for the His11 type of 2A is highlighted with colour black and blue respectively.

The result of the assay showed that none of these mutations, regardless of whether they were introduced alone or together with His11, conferred a significant activity except the single mutant L6K.These data suggested that the minor alterations were not able to compensate the activity loss caused by G11H mutation. This is consistent to the proposal that the specific combinations of amino acids throughout the N-terminal portion of 2A enable it to function (see discussion).

Another result to be noted was that the fusion 2A mutants described in last section presented slight higher activity than the separated mutants here. This suggested that even without His11, the patches selected from His11 sequence preformed better with their own contexts, which was another proof of the existence of the complex co-operation(s) within the N-terminal portion.

5.3 Summary and Discussion

The comparison of the full set of available 2A peptides revealed that besides the recognised C-terminal motif, position 11 is also conserved. 2A sequences could be classed into two types based on the amino acid (glycine or histidine) at the 11th position, and sequences of these different types presented distinct amino acids usage at the non-conserved N-terminal positions within the peptide. These findings were confirmed by a renewed alignment study with updated collection of 2A sequences in present project.

These observations, combined with the mutagenesis data obtained in chapter 3 and 4, suggest the existence of co-operation(s) between position 11 and other positions, and Gly11 or His11 has the different requirements of these positions, namely position 1 to10, and position 15. As H11G/T15S double mutations of TMEV 2A (fusion mutant T10/11F) caused dramatic loss of function, the position 15 was supported important to the 2A reaction. This view has been supported by the fact that His11 relies (at least in some contexts) on T15 (JD Brown, personal communication). The distinct 'style' within positions 1 to 10, has been confirmed as important for 2A function.

Fused 2A mutants of sequences of different type were made and tested with an *in vitro* assay. The results shows that the N-terminal half (position 1 to 9/10) and C-terminal half (position 10/11-19) from different type of 2A sequences could not function together. As the major difference of the C-terminal halves of two types of 2A sequence is the amino acid at position 11, this result is consistent with our hypothesis that Gly11 and His11 have respectively specific requirement(s) over the context within N-terminal half; and in turn, certain amino acid(s) or combination of amino acids of N-terminal half have to satisfy such requirement(s) by some way(s), to make the Gly11 or His11 functional and hence the whole sequence working.

For identifying these amino acid(s) or combination of amino acids that were interactive with amino acid at position 11, various 'snippets' picked from the N-terminal half of His11 type of 2As were introduced into the same position of FMDV 2A, having each with

Gly, or His at position 11. However, none of these sequence ‘snippets’ compensated for the loss of activity caused by the substitution of the Histidine at position 11. As the chosen snippets here covered most critical positions of the N-terminal half, ironically it seemed that only a ‘snippet’ including the whole sequence from position 1 to 10 of a His11 sequence was able to compensate mutant G11H of FMDV 2A, namely, function of position 1-11 was integrated.

Since the amino acid or combination of amino acids or the N-terminal half we picked from His11 sequences works well in their own context but do not function in the same positions of FMDV, they must fulfil their function (presumably the correct position of the C-terminal motif) co-operatively rather than each amino acid individually. Thus, these results suggest the whole N-terminal portion up to and including position 11 function as a unit, and amino acids within it are optimized to each other through the evolution.

Chapter 6: Analysis of proline 19 to stop 2A mutants-directed ribosome-nascent chain complex (19X RNC)

According to the model of 2A reaction discussed in introduction (Figure 1.5) the first two steps of the reaction promote translation termination at the Gly18 and Pro19 peptide bond formation. The peptidyl-transferase centre (PTC) of ribosome thus deviates from its normal function, and this is presumably precipitated by interaction between the nascent 2A peptide and ribosomal exit tunnel. (See section 1.3.4) A somewhat similar case of dysfunction of PTC is the ribosome stall induced by *UL4* gene uORF2, of human cytomegalovirus, which, like 2A peptide, functions only with eukaryotic ribosome.

The CMV uORF2 arrests the ribosome at the final stop codon at end of the sequence. However, it is evident that eRF1 recognize the stop codon of the uORF2 (Janzen *et al*., 2002). In the case of 2A, the prolyl-tRNA may be also able to recognise proline codon at A site while not drive peptide bond synthesis, since other tRNAs can recognize corresponding codons in the A site at end of 2A (see inactive Pro 19 mutants in section 3.3). Therefore, for both events the rearrangement(s) that are induced by the nascent chain not only prevents proper recognition of the codon in the A site but also prevents the reactions in the PTC at the same time. Hence question arises if a stop codon is put in the position 19 of 2A how does it behave?

To answer this question, prior to my study Dr.Victoria Doronina, a former postdoctoral research associate in the laboratory, made a series of FMDV 2A mutants based on the construct [pro-α-facor-2A] containing FMDV 2A with Pro19 to stop mutation (P19X). These constructs were tested in wheat germ lysate *in vitro*, to determine whether 2A peptide with a stop codon replacing P19 was, like uORF2, also able to induce a stall of the ribosome.

The result of this assay showed that in translation reactions programmed with the P19X mutant, as well as G18X, the upstream product was not released, but instead nascent chains accumulated as stable, but RNase-sensitive peptidyl-tRNA adducts. These were presumed to be ribosome-nascent chain complexes, with the 2A peptide inside the ribosomal exit tunnel (19X/18X RNC). These complexes accumulated regardless of which of the three stop codons (UAG, UAA, UGA) replaced Pro19. Collectively, the FMDV 19X/18X 2A was suggested to be able to induce ribosome stall over the stop codon as uORF2 (Sharma *et al*., 2012). The study described in this chapter focussed on this finding, and investigated it further.

The previous experiments were carried out using mRNA templates that ended at 2A, with 19X as the final codon. It was thus unknown from this work whether the same would occur, i.e. ribosomal stalling, with $2A^{P19X}$ internal to an mRNA. Further, the formation of 19X RNC was only tested with a [pro-α-factor-$2A^{P19X}$] construct and wheat germ lysate. Therefore, here I used different constructs and lysates to determine whether the complex was independent of these elements. In addition, 19X mutants for several other 2A sequences were made, to determine whether this complex was unique for the FMDV 2A sequence or a common feature for all 2A peptides.

The second part of this chapter describes confirmation that 19X RNC indeed accumulate, and steps towards purification of these complexes. Stable formation of a complex is a key prerequisite for characterisation of its structure via techniques such as cryo-electron microscopy. Finding of 19X RNC offered a good start point for such analysis and for determining such things as the conformation that 2A peptides adopt within the exit tunnel, and the effects of the peptide on ribosome structure during the 2A reaction. Therefore, attempts were made to separate the complex from *in vitro* translation reactions in which they had been generated. The stability of the 19X ribosome nascent chain complex was examined, and efforts made to improve both the construct used for in vitro translation and the purification strategy. Results of these

experiments are presented here, along with a preliminary experiment of the purification of the complex.

6.1 The formation of the 19X-RNC is independent of downstream sequence, is not affected by changing upstream context and occurs in different translation systems

While the 19X-dependent ribosome stall found by Dr.Victoria Doronina was striking, the study had some limitations. Most significant one of these was that the construct she used in these experiments was a truncated [pro-α-facor-2A] without the downstream coding region to the 2A sequence. Hence, the mutual effects of the formation of the P19X-RNC and the presence of the downstream coding region needed further study Such a study would reveal, on the one hand, whether the P19X induced ribosome stall would still occur in presence of the downstream coding region and on the other hand, what would happen to the ribosome once the pause was overcome would termination and ribosome dissociation occur, or would the ribosomes be able to re-initiate translation of downstream sequences after releasing the upstream product. Alternatively, would ribosomes read through the stop codon and continue into the downstream sequence.

6.1.1 A 19X codon directs ribosome stalling in the presence of downstream coding sequence, and is independent to the translational assay and construct

To investigate whether ribosomes stall on a 2A with a stop codon substituted for G18 or P19, as seen in the experiments with truncated mRNAs, the same set of 2A mutations examined by Dr. Doronina (in Sharma *et al*., (2010)) namely P17X, G18X, P19X, P17A, P17A+P19X) were inserted into the plasmid pLH135 encoding [CFP-2A-PAC], which was previously used in chapters 3~5, with the protocol described in section 3.2.3. The resulting pLH135 variants were then used in coupled *in vitro* transcription-translations in reticulocyte lysate, as described in section 3.2. The result of the assay is shown in

figure 6.1.

#

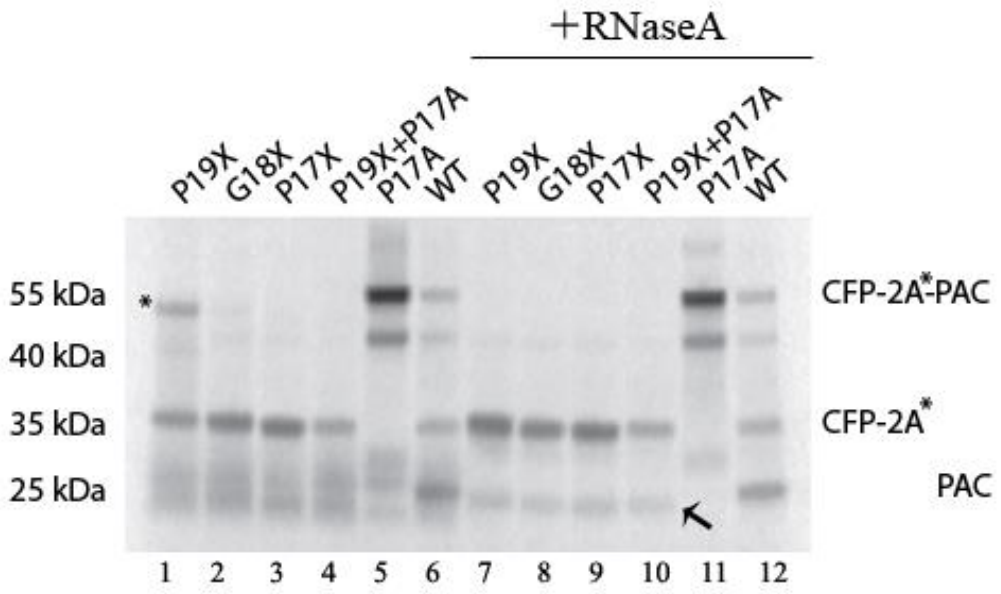


Figure 6.1 2A P19X mutant drives ribosomal stalling when downstream sequences are present. Coupled *in vitro* transcription-translation reactions were assembled in reticulocyte lysate with [S35] methionine and pLH135 containing desired 2A mutants generated by the method described in section 3.1.3. Reactions were resolved on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) without or with (+RNaseA) prior treatment with ribonuclease and visualized by phosporimager, as described in section 2.6. The peptidyl-tRNA was marked with '*'; Positions of molecular weight markers run on the gel are shown.

Results obtained with the full length [CFP-2A-PAC] polyprotein and reticulocyte lysate were consistent with the previous experiments (Sharma *et al*., 2012) with truncated construct and wheat germ, in terms of ribosome stalling: A species ~20 kDa (corresponding to the molecular weight of the tRNA) larger than the released upstream product was seen with the mutant P19X (lane 1). A similar, though faint, product was also seen with G18X (lane 2). That this represented the peptidyl-tRNA adduct was confirmed through its sensitivity to RNaseA, as the species shifted to the size of upstream product (CFP-2A) after RNaseA was added (compare lane 1,2 and lane 7,8). These data collectively suggested that during the translation of [CFP-2A-PAC] with P19X, a significant proportion of ribosomes stalled unable to release the upstream product. While the G18X mutant was also able to drive this stall, it was much less efficient in the context of further RNA sequences than when truncated mRNAs were used, most ribosomes terminating at the stop codon. Overall, the data obtained in these expeirments confirmed that the P19X and G18X-directed ribosome stall was also able to occur in presence of downstream coding region, and in the reticulocyte lysate.

As in the study of Dr. Doronina, the peptidyl-tRNA adduct was seen in neither the translation sample of P17X, (lane 3) that released upstream product as normal transcripts ended with stop codon; nor in that of P19X+P17A (lane 4), a inactive 2A mutant having a stop codon at the end. These observations suggested that the stall was the result of effects of the amino acids of 2A position 1~17, and might follow the same mechanism of 2A reaction.

In conclusion, the results presented in Figure 6.1 suggest that, the downstream coding region does not impair the P19X and G18X 2A mutant-induced ribosome stall, or at least not in a high degree; and with the stop codon at the end, 2A sequence is not able to carry on the translation of downstream product. Mostly likely, after the stall and release of the upstream product, the ribosome drops off at the final termination codon of 2A

mutant.

For further examining the universality of P19X induced translational stall to the different assay and lysate, as well as optimizing the conditions for following research, the same construct [CFP-2A-PAC] (plasmid pLH135) containing three mutants, namely P19X, P19X+P17A, and P17A, were also tested in *in vitro* translation assay of reticulocyte and wheat germ lysates, using mRNA, rather than DNA, as template.

For generating the mRNA templates, the pLH135 of these variants were linearized by Sal I, then *in vitro* transcription were carried out, following the protocol described in section 2.6.2. The resulting mRNAs were then used in *in vitro* translation assay of reticulocyte and wheat germ lysates (protocols in section 2.6.3 and 2.6.4 respectively). The result of these experiments is presented in figure 6.2

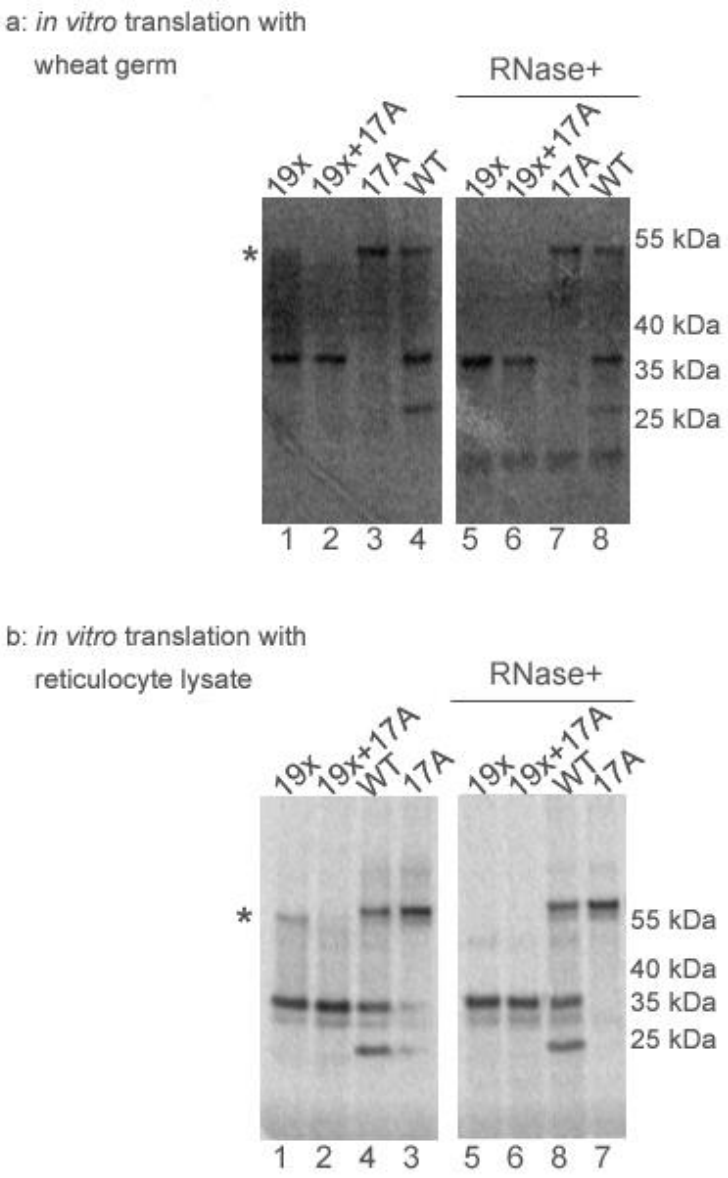


Figure 6.2 *In vitro* assay of 2A P19X and other mutants in wheat germ (a) or reticulocyte lysate (b). *In vitro* translation reactions were assembled in lysates indicated with [S35] methionine and mRNA. Reactions were resolved on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) without or with (RNase+) prior ribonuclease treatment and visualized by phosporimager, as described in section 2.6. Peptidy(2A)-tRNA was marked with '*' The positions of molecular weight markers run in the gel are indicated.

The *in vitro* experiments using mRNA as templates obtained the similar results to the coupled assay. In reticulocyte lysate, the products from the mRNAs were same as in the coupled experiments: the same RNase-sensitive adduct, representing the upstream product-tRNA adduct, was observed in the same mutant (19X), at the identical size, and in the similar proportion with the released upstream products (compare figure 6.2b lane 1 and the figure 6.1 lane 1). Other three mutants also behave as the previous experiment: the 17A+19X mutant generated no adduct, while wild type 2A and 17A inactive 2A mutant presented the separated and full length products respectively. On the other hands, the assay with the wheat germ lysate also gave the same result. Here, the 19X mutant indeed produced the adduct, which was, in the experiment presented degraded slightly (6.2a. lane 1). In these experiments using mRNA to program the translation reactions, no 25 kDa species was seen in when 2A mutants were used (compare with Fig 6.1). The same set of mRNAs were also tested in the third lysate used in this chapter, the yeast extract. However, in this case, the *in vitro* assay failed to generate any detectable signal (Data not shown).

In summary, the data obtained is this section combined with that of Dr. Victoria Doronina, (truncate construct [polyhis pro-α-facor-2A], with wheat germ lysate) suggest that 19X-depentent ribosome arresting and is independent to the construct, lysate and template.

6.1.2 Ribosome stalling directed by P19X occurs with different 2A peptides

The study of P19X dependent ribosome arresting carried out so far was all based on the mutants of 2A sequence from FMDV. As discussed in the chapters of mutagenesis study, the different 2A sequences, especially the positions 1-11, may adopt different strategies to function. It was therefore important to ask whether a ribosomal stall at a P19X mutation could also occur in other 2A peptides besides FMDV. Hence, in this section, the mutation P19X was made to other 2A sequences to test its ability to arrest the ribosome in alternative 2A contexts.

Three 2A sequences besides FMDV were selected for this study. These were from Theiler's murine encephalomyelitis virus (TMEV), infectious myonecrosis virus (IMNV) and pseudorabies virus (Prv-1) TMEV 2A is a representative of His11 type of 2As, while the IMNV and Prv-1 2A peptides have very different amino acid composition to FMDV 2A, and do not display the strong α-helical propensity within their N-terminal portion. All three 2A sequences have high activity in previous study using the [GFP-2A-GUS] construct (Donnelly *et al*., 2001). The version of TMEV 2A contained a M15S mutation, which, as discussed in chapter 5, should be a neutral change that not affects the 2A function. The sequences of these 2A peptides were presented in table 6.1 column ‘sequence’.

Host virus	Sequence
FMDV(start plasmid)	TLNFDLLKLAGDVESNPGP
TMEV(S15M)	HADYYKQRLIHDVESNPGP
Prv-1	VGGRGSLLTCGDVESNPGP
IMNV	MLPPPDLTSCGDVESNPGP

Table 6.1 the sequence of 2A peptide tested in this section

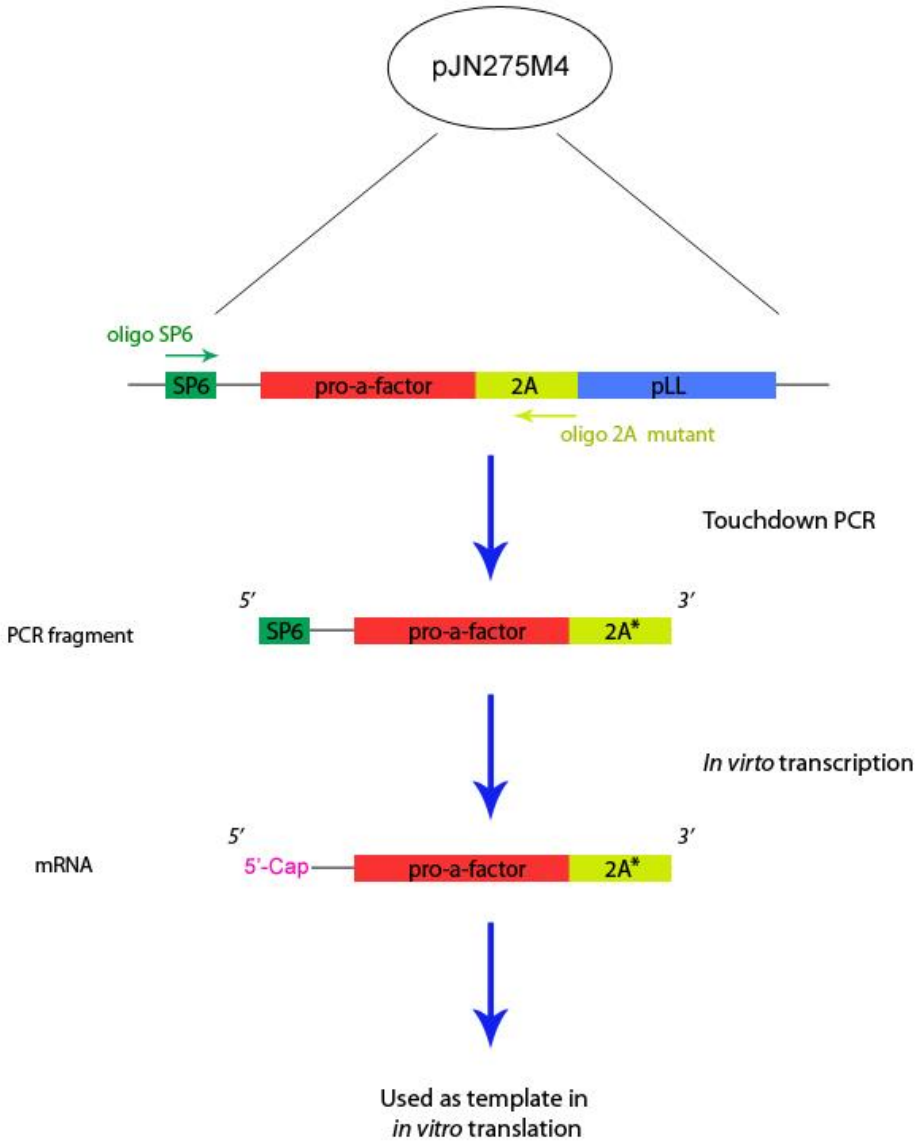
The selected 2A sequences were generated and inserted into plasmid pJN275M4 via the method described in section 5.2.1. pJN275M4 is a variant of pJN141 that Dr. Doronina used in previous study (Sharma *et al*., 2012) (see section 6.2.1), and contains [pro-α-factor-2A-prolactin] preceeded by a 8xHIS-tag.

The three resulting plasmids, after linearized by restriction enzyme PvuII, were transcribed with SP6 *in vitro*, and the resulting mRNAs were tested in *in vitro* translation assay using wheat exacts. The samples of the translation were resolved on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) and visualized by phosporimager (protocol see section 2.6). All the selected 2A sequences, including the M15S-mutated TMEV, gave high level of 2A activities with the reporter, and were therefore characterized as functional 2A sequence (Data not shown). This also conforms the previous finding and extends it to the different context, i.e. of the alpha factor fusions.

To examine whether P19X mutants of three selected 2A sequences were able to induce ribosomal stalling, two truncated constructs containing wild type sequence and mutant P19X ([polyhis tag-pro-α-facor-2A] and [polyhis tag-pro-α-facor-2A^{P19X}]) were generated from the constructs containing each one of these 2A peptides, along with the FMDV peptide as a positive control. The strategy used here to generate truncated constructs was one of the basic methods of the study described in this chapter and used excessively in the research described in section 6.2. Hence, a more detailed introduction on this method is presented here:

For making the mutants, the oligonucleotides comprising the reverse 2A sequence with desired changes (in this case, P19X or wild type) were designed. Then, touchdown PCR was performed, as described in Methods and Materials section 2.4.3, with the above oligonucleotide and an SP6 promoter primer, and plasmid pJN275M4, that contains SP6 promoter followed by the open reading frame, as template. Thus, the resulting PCR fragment of coding sequence of [polyhis tag-pro-α-facor-19X/WT] was generated. Using this fragment as template, *in vitro* transcription with 5'cap analog was applied, as described in section 2.6.2. Finally the 5'-capped mRNA of the construct

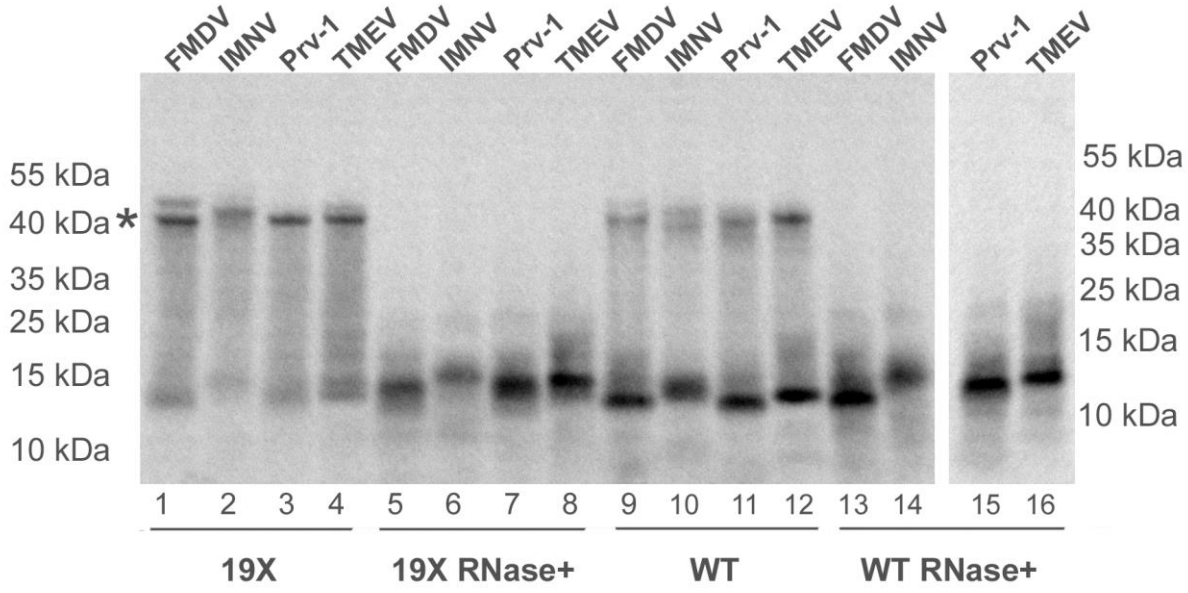
[pro-α-facor-desired 2A mutant] is obtained, and used in following *in vitro* translation test with various lysate systems. (Figure 6.3)



.

Figure 6.3 General strategy of the generation of truncated 2A mutants. See text for details

The mRNA ended with P19X mutant and wild type of the three selected 2A peptides and FMDV 2A were generated as the strategy described above, and tested by *in vitro* assay of wheat germ lysate. The result of the assay is shown in figure 6.4



.

Figure 6.4 The P19X mutation directs ribosome stalling in the context of different 2A sequences.Truncated mRNA was prepared from plasmid pJN275M4 variants containing 2A sequences from FMDV, IMNV, Prv-1 and TMEV. *In vitro* translation reactions were assembled in wheat germ with [S35] methionine and the mRNA. Reactions were resolved on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) without or with (RNase+) prior treatment with ribonuclease, and visualized by phosporimager, as described in section 2.6. Peptidy(2A)-tRNA was marked with '*' Positions of molecular weight markers run on the gel are shown.

Results obtained from different 2A peptides were very similar. The adducts ~40 kDa representing the peptidy-tRNA presented with the P19X mutants based on all four 2A sequences (lanes 1 to 4). As expected, these adducts were RNaseA-sensitive. (compare lanes 1 to 4 and 5 to 8) On the other hand, in the sample of all 4 wild type 2A sequence, the majority of translation yield was the upstream products with the predicted size of expected from their sequence (lanes 9 to 12). Notably, the ~40 kDa, RNaseA-sensitive adducts, albeit faint, were also seen with wild type 2A sequences. This species was observed in the previous study (Sharma *et al*., 2012). This can be explained as the representative of the ribosome stall induced by non-stop open reading frame: of wild type 2As the separation activities never reached 100 %, which means that small proportion of ribosomes are passing the 2A sequence and continuing translate to generate full length product rather than 'trapped' into the 2A reaction; similarly, during the translation of the truncated construct ended with wild type 2A, minority of ribosome 'escaped' from the 2A mechanism, and as the construct has not stop codon, they are stalled by the non-stop mRNA. An argument in favour of this explanation is that the amount of the ~40 kDa, RNaseA-sensitive adducts in the sample of a wild type 2A, is roughly inversely proportional to the separation activity of the 2A.

These observations agree with the argument that, the P19X mutation induced ribosome stalling are the same as FMDV 2A with other three 2A sequences, and, though I have not tested an exhaustive set of 2A peptides with P19X, the data obtained with the 4 peptides, which have very varied amino acid sequence, suggest that it may be a phenomenon that would occur with all 2A peptides.

6.2 Purification of the 19X directed ribosome nascent chain complex

As shown in the section 1.2.4 (introduction), in studies of translation-arrested peptides, cryo-EM has generated some of the most insightful data in recent years. 2A peptides also pause the translation during 2A reaction (section 1.3), but this pause is very short (Doronina *et al*., 2008), making it infeasible to investigate by cryo-EM, which requires the purified ribosome particle with the nascent peptide inside its exit tunnel. The P19X mutant provides, for the first time, a substrate that may be sufficiently stable to purify for structural analysis. This section describes the generation, optimization and a preliminary purification of 19x RNC with the ribosomes stalled on a P19X 2A mutant.

6.2.1 Construct of the template for the purification

The construct selected to be used in purification experiments was based on the truncated [pro-α-facor-2A-pLL] with 2A P19X, that was similar to the one used in section 6.1.2 and previous study. This provided a nascent chain length of ~70 amino acids, which guarantees that when P19X-induced translation stall occurs, the N-terminal end of the nascent chain has emerged from the ribosomal exit tunnel (which can hold the length of ~40 amino acids in compacted form). Thus, a 'tag' suitable for purification can be added to the N-terminal end. On the other hands, though the effect of the length of the nascent chain to the stall complex is largely unclear, any excessive amino acids on the peptide chain might impair the stability of the complex. The use of the truncated construct ensured that all the stalled ribosomes would be at P19X – ribosomes either terminate or stall at this mutation (See Figure 6.1 and 6.2).

Prior to my study, a poly-His tag was added to the N-terminal end of construct [pro-α-facor-2A(FMDV)-pLL] in the plasmid pJN141, which Dr. Doronina used in previous study (Sharma *et al*., 2012), for the purification of P19X RNC using his-tag affinity Ni resin. I generated His-tagged truncated construct [pro-α-facor-2A] from the resulting plasmid (pJN275) with the method described in section 6.1.2, and examined it *in vitro*. However, besides the expected species ~12 kD representing the translation product, (Figure 6.5 a lane 2, marked with '*') a prominent ~9 kD species was also seen when the reaction was treated with RNaseA (Figure 6.5 a lane 2, marked with red'*'). Given its size, and the position of methionine codons in the coding sequence of the mRNA, this was suspected to be the product of translation in which the ribosome 'skipped' the start codon of the mRNA, started at the second methionine to the N-terminal end. Indeed, in plasmid pJN275, the start codon is very close to the SP6 promoter, i.e. 5'-end of the mRNA. Such a short distance can cause ribosomes to miss the start codon. As the poly-His tag was added just after the first methionine codon, skipping of the start codon meant a large proportion of constructs translated would not contain the tag.

To overcome this problem, a spacer 80 nucleotides sequence was inserted (through EcoRI) between the SP6 promoter and the start of the open reading frame, (using the strategy described in section 5.2.1). This space both extended the distance of start codon to the 5'-end of mRNA, but also placed the start codon the Kozak consensus sequence (GCCACC**ATG**G) that facilitate efficient initiation. Translation of mRNA from the resulting plasmid (pJN275M4), (Figure 6.5B) revealed the majority translation product (after RNaseA treatment) to be 12 kDa. The conclusion reached was that the the start codon was being efficiently recognised, and therefore the plasmid pJN275M4 was selected as the starting construct for further experiments.

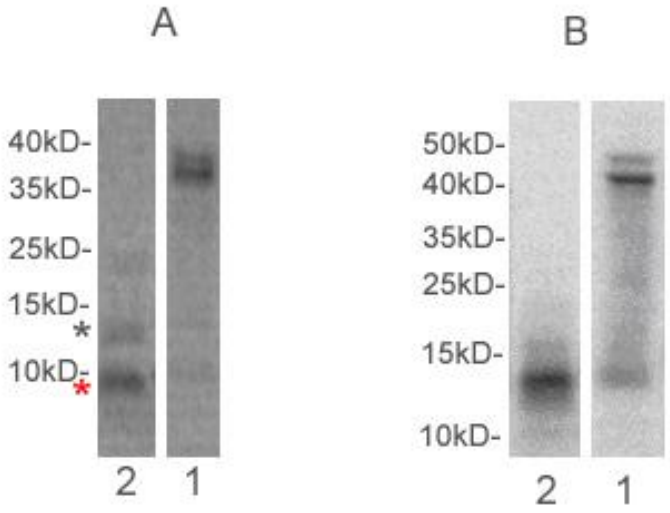


Figure 6.5 The optimization of construct [pro-α-facor-2A]. Truncated mRNA containing FMDV P19X mutant was prepared from plasmid pJN275 and pJN275M4 containing 2A *In vitro* translation reactions were assembled in wheat germ with [S35] methionine and the mRNA. Reactions were resolved on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) without (1) or with (2) prior treatment with ribonuclease, and visualized by phosporimager, as described in section 2.6. Positions of molecular weight markers run on the gel are shown. See text for detail

Besides the construct, wheat germ lysate was chosen as the best system to generate the P19X RNC. In parallel experiments this provided the best yield and stability of the complex (data not shown)

6.2.2 Preliminary purification of the P19X RNC

After generation and testing of plasmid templates and selection of the lysate, a preliminary purification experiment was carried out. This experiment followed a protocol adapted from (Halic *et al*., 2004), and which is described in the Materials and Methods (section 2.7). There were two main steps in this method following the translation reaction itself. The first step is sucrose gradient centrifugation, in which ribosomes (including RNCs), are centrifuged into a pellet thereby separating them from the rest of the sample. The pellet is then resuspended in buffer and the second purification step is immobilised metal ion affinity chromatography on Ni-NTA agarose. This affinity purification step should specifically isolate the RNC carrying the His-tagged RNC from, other ribosomes as well as other pelleted material.(see section 2.7 for details)

.

Three truncated 2A variants were selected as the subjects for a prelimianry purification experiment: FMDV-P19X and FMDV wild type, both generated from pJN275M4, with the wild type sequence and FMDV-P19X generated from pJN141, which contained untagged [pro-α-facor-2A-pLL]. The wild type 2A acted as a control that released the nascent chain, which should then not be present in the pellet of sucrose gradient centrifugation. The untagged P19X product should remain ribosome associated, but not purify on the Ni-NTA resin. Truncated mRNAs were generated from the templates amplified from respective plasmids, with the method described in section 6.1.2. These mRNAs were then assembled with the wheat germ extracts and [S35] methionine for *in vitro* translation, and the centrifugation and affinity purification steps followed. After purification, the beads used to bind the complex were heated in loading dye, at 80℃ for 5 mins to release any material remaining associated with the beads following the elution steps, Fractions derived from each step of the protocol were then separated on a Nu-PAGE 4-12% Bis-Tris gel (Invitrogen) and visualized by phosporimager. The result of this experiment is presented in figure 6.6

Figure 6.6 a the sucross centrifuge and binding supernate

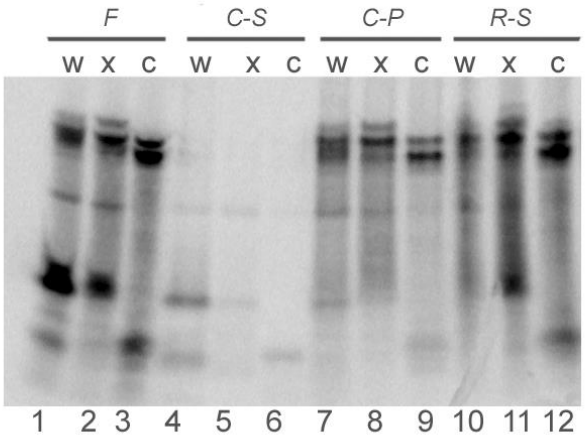


Figure 6.6 b poly His-tag purification

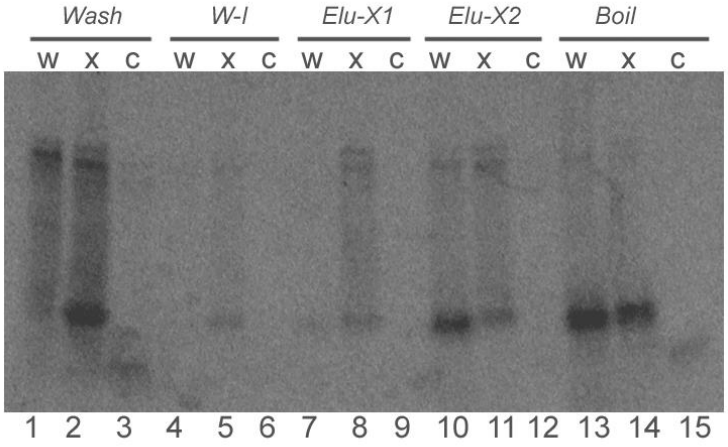


Figure 6.6 Purification of the 19x RNC. w, x, c refer to FMDV-wt, FMDV-19x (both His-tagged), and FMDV-19x from pJN141 respectively, see text for details a. the sucrose centrifuge: the *F, C-S, C-P, R-S* refers to full reaction, centrifuge-supernate, resuspended centrifuge-pellet and binding supernate respectively. b. beads of poly His-tag purification: *Wash, W-I, Elu-X1, Elu-X2,* refer to the elution from the beads washed with buffer containing 250 and 500 potassium acetate, 10mM imidazole, 100mM imidazole, 200mM imidazole, respectively. '*Boil'* refer to material released from the beads following heating. See text and section 2.7 for details. The peptidy-tRNA adduct representing the 19X-RNC, is marked with '*'

The result of sucrose gradient centrifugation is generally positive. In the *in vitro* translation, the 2A mutants generated the products as expected. The two P19X mutants gave a relatively low proportion of released peptide, (a. lanes 2 and 3) compared to the wild type 2A where most of the nascent chains were released. (a. lane 1) Peptidyl-tRNAs were clearly isolated in the pellet following sucrose gradient centrifugation, as there was no adducts representing the P19X-RNC (or non-stop mRNA induced RNC in the case of wild type 2A, see section 6.1.2) can be seen only in the resuspended pellet, in which very small amount of released peptide detected; (a. lanes 7~9) on contrast no adduct but only peptides can be found in supernate.(a. lanes 4~6)

Subsequent purification on Ni-NTA resin was less successful. The similarity of the amount of P19X RNC in supernate between before and after the Ni-NTA beads binding in two his-tagged mutant, (w,x) indicated low binding efficiency; (compare a. lanes 7,8 with 10,11) and during and after the binding, a high proportion of the nascent chains were released from the complexes (or degraded), implying that the binding condition itself was not favour able to the stability of the complex. Indeed, after removal of the unspecific binding by wash (b. lanes 1~6), only a very small amount of intact RNCs were eluted from the beads (lanes 7~12). However, this was specific, as the specific elution was only detected with two his-tagged mutants (b. lanes 7,8,10,11) but not the untagged control. (b. lanes 9,12)

6.3 Summary and discussion

The study discussed in this chapter extends the finding of FMDV 2A mutants P19X and G18X-induced ribosome stall to different translation systems and constructs. The evidence that this stall could occur in presence of downstream coding region is also provided (section 6.1.1, Figure 6.1). However comparison with Dr. Doronina's data revealed that a larger proportion of the upstream product was seen with the P19X in the construct [GFP-2A-PAC] than the same mutant in truncated constructs [pro-α-facor-2A]. This indicated here the P19X-RNC is less stable than the previous study. As in the wheat germ lysate used in previous study, like reticulocyte lysate, also gave the high level of released peptide with P19X (Figure 6.2), this can not be a result of the different lysates. Therefore, there are only two possible reasons for the lower stability of the complex in present experiment:

1) The presence of mRNA sequence downstream to the 2A does influence the P19X-induced ribosome stalling. Although the stalling still occurs, the length of the stall is shorted.

2) The stability of P19X-RNC is also affected by the upstream context.

On the other hand, the presenting the stop codon in the final position of 2A sequence has been proved being able to remove the re-initiation of 2A of the translation. This confirmed that the two main events of 2A function, the non-stop codon releasing and the re-initation of translation are coupled, rather independent process.

Chapter7 Summary and discussion

The 2A reaction comprises two main unconventional events during translation: ribosome releasing the upstream product at end of the sequence without a stop codon, and re-initiation of the translation of downstream coding region, with Pro19 as the start codon. While the mechanism of the re-initiation is almost totally unknown (though it is generally believed coupled with the first event), a model, suggested by the current research, has been proposed to explain the non stop codon releasing. According to this model, the amino acids located at position 1~17th of 2A sequence, through interaction with the ribosomal components, directs the peptidyl transferase centre of ribosome deviating its normal function, thus, while the PTC can still catalyse the peptide bond formation for other amino acids pairs, it is not able to generate the bond between proline and glycine at the 18th and 19th position of 2A, which is the most demanding amino acids pair for peptide bond formation. Therefore, the study described in this thesis including two themes:

1) Site-specific mutagenesis study of 2A sequence, for investigating the functional elements within 2A sequence and detecting the potential interaction between the 2A nascent chain and ribosomal exit tunnel;

2) Analysis of the finding of 2A peptides with Pro19 and Gly18 to stop codon mutation, which direct the ribosome stalling during the translation. This not only reveals, potentially due to the effect of position 1~17 of 2A, the dysfunction of PTC of ribosome, but also give a profound perspective for the further cryo-EM study.

Originally as a complimentary study to a previous 2A random mutant screen, the site-specific mutagenesis was carried out to the whole sequence of 2A. The mutants, except ones that largely deviated from the sequence of FMDV 2A, was generated by the

site-specific PCR described in section 3.1. In present study, the separation activity of 2A mutant were examined by two independent reporter system: an *in vivo* assay that exploited the varied colour (white to red) of yeast expressing reducing amounts of Ade2p and an *in vitro* translation assay using the poly-protein construct [GFP-2A-PAC] (pLH135). The *in vivo* assay for Val13 and Gly11 of 2A sequence filled the 'holes' of the previous random screen. The result found that all mutations at these two positions reduce the 2A activity dramatically, and therefore there are strict constrains for the 13 and 11 position. This result, combined with previous data obtained from the random screen, suggests the whole 19 amino acid in 2A sequence, despite the variability of the N-terminal portion, is less flexible than expected. Thus, unlike many other peptides manipulating the translation process, the amino acids in the whole 2A sequence participate the reaction in some way.

Many of the 2A sequences, including the one from FMDV, have strong α-helical propensity on their N-terminal portion. For examining whether the secondary structure α-helix was critical for the 2A function, as well as addressing the general duration pattern of the positions in 2A sequence against mutations, three amino acids, alanine, proline and glycine, that have the different α-helical propensity, were systematically substituted into each position of FMDV 2A sequence, then tested by coupled *in vitro* assay in reticulocyte lysate. Among three sets of mutants, similar trend was observed: the closer a position is to the C-terminal end, the more serious impairment that results from the substitution, with only one major exception of position 15. This result coincides with previous 2A alignment study, suggest that the positions close to the N-terminal end are less demanding than that of C-terminal end. Indeed, among all 11 mutations at first two positions to the C-terminal end examined by present study and previous random screen, only 2 of them (L1P, L1Q) showed obvious activity decrease. Hence, the contribution of amino acids at position 1 and 2 may be simply excluding certain unfavorable amino acids and setting good context for the 2A peptide. On the other hand, as the distinct secondary structure propensity of proline glycine and alanine, the resemblance of the general trend presented by the systematical replacement screen with this three amino acid implies that α-helical propensity previously thought critical

for the 2A function might not be crucial, or at least that the structure of the backbone alone could not satisfy the requirement(s) of 2A function at most positions, other factors such as side chain must be considered.

The abundant information collected from the mutagenesis study, as well as structural studies of other peptides manipulating the translational process allows us to estimate the potential interactions between nascent chain of 2A and the ribosomal component within the exit tunnel. The preliminary estimation is discussed in this final section. The structural study using cryo-EM method of two eukaryotic ribosomal arresting peptides, Cytomegalovirus UL4 uORF2 and arginine attenuator peptide (AAP), reveals that the both peptides contact the ribosomal proteins L4/L17 emerging from wall of the constriction region of the exit tunnel, with their 10th (Ser12 of uORF2) and 12th (Asn12 of AAP) amino acids counted from PTC. The mutagenesis studies of uORF2 and AAP, as well as many other structural and biochemical studies on ribosome stalling peptides of prokaryotic translation, confirm such contacts critical for the function of these short peptides, and that they are the most common interactions that the peptides driving translation may apply. According to corresponding length of amino acids chain of 2A sequence to that of uORF2 and AAP, the 2A peptide, if it also interacts with the rprotein in the exit tunnel constrain region, it may interact with these rproteins with its amino acid(s) located at position 6~9. Since during the reaction the both of uORF2 and AAP adopt compact form within the exit tunnel, providing the 2A adopt a looser conformation, the amino acid interacting with the ribosomal components may also be the one at position 11, the importance of which has been proved by the mutagenesis study presented in this study. If it is the case, it will be a reasonable explanation to the finding that the 2A peptides with glycine or histidine at position 11 having distinct strategy to drive 2A reaction.

The structural study of AAP and uORF2 also reveals the extensive interactions of the C-terminal half of these short peptides with the nucleotides of the wall of the upper tunnel region of the exit tunnel. This might also happen during the 2A reaction. While the mutagenesis study suggests in AAP and uORF2 some of these contacts may not be

critical, for the 2A reaction the potential contacts between the C-terminal portion and rRNA should be important, as the data presented in this and previous study shows that any change on the sequence of the conserved C-terminal portion, except position 15, dramatically impairs, or abolishes the separation activity.

During the 2A reaction, the N-terminal portion of 2A sequence mainly rests in the lower tunnel region of the exit tunnel. A possible explanation of its function is to correctly position the conserved C-terminal portion including position 11, for facilitating their potential interactions with ribosomal component.

The second theme of the study presented in this thesis is the analysis of the P19X and P18X 2A mutants-induced translation stalling. A previous mechanism research revealed that, a construct ending with FMDV 2A mutants P19X and G18X directed the ribosome stalling during the translation process, forming the so called P19X/G18X directed ribosome nascent chain complex (P19X/G18X RNC). This previous research used truncated construct with no downstream coding region to 2A. Here I found that the stall still occurred, though in a compromised level, in presence of the downstream sequence; and that with the P19X/G18X, the translation of downstream sequence to 2A is failed to be re-initiated. The P19X-induced ribosome stalling was also examinied with different construct and *in vitro* translation lysates, and 2A peptides from different origins. The results showed this stall, though in variable levels, occurred in all conditions that were been tested. These observations extend this finding, suggesting that the stalling is independent to the construct, lysate used in the test, and is universal among all the functional 2A peptides.

The P19X/G18X-induced ribosome stalling offers further evidence that the sequence of the position 1~17 of 2A directs the ribosomal peptidyl-transferase centre into dysfunction, presumably through the potential interactions between the nascent chain and ribosomal components. Moreover, it also provides a profound perspective for

further structural analysis. During the 2A reaction there is indeed a short pause, but this is too short for cryo-EM study. Here providing that the wild type 2A and P19X/G18X-induced ribosome stalling follows same mechanism to disturb the function of the PTC, the structural insight of 2A peptide can be obtained through the cryo-EM analysis for P19X-RNC, which is stable enough to be tested by the technique. For this purpose, a construct had been built (pJN275M4) and preliminary purification of the P19X-RNC was preformed. However, the experimental conditions need further optimization to meet the requirement of cryo-EM study.

Reference

ACKER, M. G. AND J. R. LORSCH (2008). "MECHANISM OF RIBOSOMAL SUBUNIT JOINING DURING EUKARYOTIC TRANSLATION INITIATION." BIOCHEM SOC TRANS **36**(PT 4): 653-657.

ALKALAEVA, E. Z., A. V. PISAREV, ET AL. (2006). "IN VITRO RECONSTITUTION OF EUKARYOTIC TRANSLATION REVEALS COOPERATIVITY BETWEEN RELEASE FACTORS ERF1 AND ERF3." CELL **125**(6): 1125-1136.

AMRANI, N., S. GHOSH, ET AL. (2008). "TRANSLATION FACTORS PROMOTE THE FORMATION OF TWO STATES OF THE CLOSED-LOOP MRNP." NATURE **453**(7199): 1276-1280.

BEIER, H. AND M. GRIMM (2001). "MISREADING OF TERMINATION CODONS IN EUKARYOTES BY NATURAL NONSENSE SUPPRESSOR TRNAS." NUCLEIC ACIDS RES **29**(23): 4767-4782.

BEN-SHEM, A., N. GARREAU DE LOUBRESSE, ET AL. "THE STRUCTURE OF THE EUKARYOTIC RIBOSOME AT 3.0 A RESOLUTION." SCIENCE **334**(6062): 1524-1529.

BERINGER, M. AND M. V. RODNINA (2007). "THE RIBOSOMAL PEPTIDYL TRANSFERASE." MOL CELL **26**(3): 311-321.

BERK, V., W. ZHANG, ET AL. (2006). "STRUCTURAL BASIS FOR MRNA AND TRNA POSITIONING ON THE RIBOSOME." PROC NATL ACAD SCI U S A **103**(43): 15830-15834.

BHUSHAN, S., H. MEYER, ET AL. "STRUCTURAL BASIS FOR TRANSLATIONAL STALLING BY HUMAN CYTOMEGALOVIRUS AND FUNGAL ARGININE ATTENUATOR PEPTIDE." MOL CELL **40**(1): 138-146.

BHUSHAN, S., ET AL., SECM-STALLED RIBOSOMES ADOPT AN ALTERED GEOMETRY AT THE PEPTIDYL TRANSFERASE CENTER. PLOS BIOL. 9(1): P. E1000581.

BRAKIER-GINGRAS, L., J. CHARBONNEAU, ET AL. "TARGETING FRAMESHIFTING IN THE HUMAN IMMUNODEFICIENCY VIRUS." EXPERT OPIN THER TARGETS **16**(3): 249-258.

BRIERLEY, I., P. DIGARD, ET AL. (1989). "CHARACTERIZATION OF AN EFFICIENT CORONAVIRUS RIBOSOMAL FRAMESHIFTING SIGNAL: REQUIREMENT FOR AN RNA PSEUDOKNOT." CELL **57**(4): 537-547.

BRIERLEY, I. AND F. J. DOS RAMOS (2006). "PROGRAMMED RIBOSOMAL FRAMESHIFTING IN HIV-1 AND THE SARS-COV." VIRUS RES **119**(1): 29-42.

BRIERLEY, I., S. PENNELL, ET AL. (2007). "VIRAL RNA PSEUDOKNOTS: VERSATILE MOTIFS IN GENE EXPRESSION AND REPLICATION." NAT REV MICROBIOL **5**(8): 598-610.

CAO, J. AND A. P. GEBALLE (1995). "TRANSLATIONAL INHIBITION BY A HUMAN CYTOMEGALOVIRUS

upstream open reading frame despite inefficient utilization of its AUG codon." J Virol **69**(2): 1030-1036.

Cao, J. and A. P. Geballe (1996). "Coding sequence-dependent ribosomal arrest at termination of translation." Mol Cell Biol **16**(2): 603-608.

Cao, J. and A. P. Geballe (1996). "Inhibition of nascent-peptide release at translation termination." Mol Cell Biol **16**(12): 7109-7114.

Carter, A.P., et al., Crystal structure of an initiation factor bound to the 30S ribosomal subunit. Science, 2001. 291(5503): p. 498-501.

Chavatte, L., et al., The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. Embo J, 2002. 21(19): p. 5302-11.

Cenik, C., H. N. Chua, et al. "Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes." PLoS Genet **7**(4): e1001366.

Cheng, Z., K. Saito, et al. (2009). "Structural insights into eRF3 and stop codon recognition by eRF1." Genes Dev **23**(9): 1106-1118.

Chou, P.Y. and G.D. Fasman, Empirical predictions of protein conformation. Annu Rev Biochem, 1978. 47: p. 251-76.

Cogoni, C. and G. Macino (2000). "Post-transcriptional gene silencing across kingdoms." Curr Opin Genet Dev **10**(6): 638-643.

Craigen, W. J. and C. T. Caskey (1987). "The function, structure and regulation of E. coli peptide chain release factors." Biochimie **69**(10): 1031-1041.

Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.

Cruz-Vera, L. R., S. Rajagopal, et al. (2005). "Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression." Mol Cell **19**(3): 333-343.

Cruz-Vera, L. R. and C. Yanofsky (2008). "Conserved residues Asp16 and Pro24 of TnaC-tRNAPro participate in tryptophan induction of Tna operon expression." J Bacteriol **190**(14): 4791-4797.

de Felipe, P. (2002). "Polycistronic viral vectors." Curr Gene Ther **2**(3): 355-378.

de Felipe, P. (2004). "Skipping the co-expression problem: the new 2A "CHYSEL"

technology." Genet Vaccines Ther **2**(1): 13.

De Felipe, P. and M. Izquierdo (2000). "Tricistronic and tetracistronic retroviral vectors for gene transfer." Hum Gene Ther **11**(13): 1921-1931.

de Felipe, P. and M. Izquierdo (2003). "Construction and characterization of pentacistronic retrovirus vectors." J Gen Virol **84**(Pt 5): 1281-1285.

de Felipe, P., V. Martin, et al. (1999). "Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy." Gene Ther **6**(2): 198-208.

Demeshkina, N., et al., A new understanding of the decoding principle on the ribosome. Nature. 484(7393): p. 256-9.

Dieci, G., G. Fiorino, et al. (2007). "The expanding RNA polymerase III transcriptome." Trends Genet **23**(12): 614-622.

Dinman, J. D. "Mechanisms and implications of programmed translational frameshifting." Wiley Interdiscip Rev RNA **3**(5): 661-673.

Donnelly, M. L., D. Gani, et al. (1997). "The cleavage activities of aphthovirus and cardiovirus 2A proteins." J Gen Virol **78 (Pt 1)**: 13-21.

Donnelly, M. L., L. E. Hughes, et al. (2001). "The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences." J Gen Virol **82**(Pt 5): 1027-1041.(A)

Donnelly, M. L., G. Luke, et al. (2001). "Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'." J Gen Virol **82**(Pt 5): 1013-1025. (B)

Doronina, V. A., P. de Felipe, et al. (2008). "Dissection of a co-translational nascent chain separation event." Biochem Soc Trans **36**(Pt 4): 712-716.(A)

Doronina, V. A., C. Wu, et al. (2008). "Site-specific release of nascent chains from ribosomes at a sense codon." Mol Cell Biol **28**(13): 4227-4239.(B)

Dos Ramos, F., M. Carrasco, et al. (2004). "Programmed -1 ribosomal frameshifting in the SARS coronavirus." Biochem Soc Trans **32**(Pt 6): 1081-1083.

Firth, A. E., N. M. Wills, et al. "Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element." Nucleic Acids Res **39**(15): 6679-6691.

FANG, P., ET AL., A NASCENT POLYPEPTIDE DOMAIN THAT CAN REGULATE TRANSLATION ELONGATION. PROC NATL ACAD SCI U S A, 2004. 101(12): P. 4059-64.

FANG, P., Z. WANG, AND M.S. SACHS, EVOLUTIONARILY CONSERVED FEATURES OF THE ARGININE ATTENUATOR PEPTIDE PROVIDE THE NECESSARY REQUIREMENTS FOR ITS FUNCTION IN TRANSLATIONAL REGULATION. J BIOL CHEM, 2000. 275(35): P. 26710-9.

FROLOVA, L., A. SEIT-NEBI, AND L. KISSELEV, HIGHLY CONSERVED NIKS TETRAPEPTIDE IS FUNCTIONALLY ESSENTIAL IN EUKARYOTIC TRANSLATION TERMINATION FACTOR ERF1. RNA, 2002. 8(2): P. 129-36.

GARZA-SANCHEZ, F., B. D. JANSSEN, ET AL. (2006). "PROLYL-TRNA(PRO) IN THE A-SITE OF SECM-ARRESTED RIBOSOMES INHIBITS THE RECRUITMENT OF TRANSFER-MESSENGER RNA." J BIOL CHEM **281**(45): 34258-34268.

GIEDROC, D. P. AND P. V. CORNISH (2009). "FRAMESHIFTING RNA PSEUDOKNOTS: STRUCTURE AND MECHANISM." VIRUS RES **139**(2): 193-208.

GONG, F., K. ITO, ET AL. (2001). "THE MECHANISM OF TRYPTOPHAN INDUCTION OF TRYPTOPHANASE OPERON EXPRESSION: TRYPTOPHAN INHIBITS RELEASE FACTOR-MEDIATED CLEAVAGE OF TNAC-PEPTIDYL-TRNA(PRO)." PROC NATL ACAD SCI U S A **98**(16): 8997-9001.

GONG, F. AND C. YANOFSKY (2002). "INSTRUCTION OF TRANSLATING RIBOSOME BY NASCENT PEPTIDE." SCIENCE **297**(5588): 1864-1867.

HAHN, H. AND A.C. PALMENBERG, DELETION MAPPING OF THE ENCEPHALOMYOCARDITIS VIRUS PRIMARY CLEAVAGE SITE. J VIROL, 2001. 75(15): P. 7215-8.

HARRELL, L., U. MELCHER, ET AL. (2002). "PREDOMINANCE OF SIX DIFFERENT HEXANUCLEOTIDE RECODING SIGNALS 3' OF READ-THROUGH STOP CODONS." NUCLEIC ACIDS RES **30**(9): 2011-2017.

HALIC, M., ET AL., STRUCTURE OF THE SIGNAL RECOGNITION PARTICLE INTERACTING WITH THE ELONGATION-ARRESTED RIBOSOME. NATURE, 2004. 427(6977): P. 808-14.

HELM, M. (2006). "POST-TRANSCRIPTIONAL NUCLEOTIDE MODIFICATION AND ALTERNATIVE FOLDING OF RNA." NUCLEIC ACIDS RES **34**(2): 721-733.

HERAS, S. R., M. C. THOMAS, ET AL. (2006). "L1TC NON-LTR RETROTRANSPOSONS FROM TRYPANOSOMA CRUZI CONTAIN A FUNCTIONAL VIRAL-LIKE SELF-CLEAVING 2A SEQUENCE IN FRAME WITH THE ACTIVE PROTEINS THEY ENCODE." CELL MOL LIFE SCI **63**(12): 1449-1460.

HOFSTETTER, H., H. J. MONSTEIN, ET AL. (1974). "THE READTHROUGH PROTEIN A1 IS ESSENTIAL FOR THE FORMATION OF VIABLE Q BETA PARTICLES." BIOCHIM BIOPHYS ACTA **374**(2): 238-251.

HOLCIK, M., N. SONENBERG, ET AL. (2000). "INTERNAL RIBOSOME INITIATION OF TRANSLATION AND

THE CONTROL OF CELL DEATH." TRENDS GENET **16**(10): 469-473.

ITO, K. AND S. CHIBA "ARREST PEPTIDES: CIS-ACTING MODULATORS OF TRANSLATION." ANNU REV BIOCHEM **82**: 171-202.

IVANOV, I. P., C. B. ANDERSON, ET AL. (2004). "IDENTIFICATION OF A NEW ANTIZYME MRNA +1 FRAMESHIFTING STIMULATORY PSEUDOKNOT IN A SUBSET OF DIVERSE INVERTEBRATES AND ITS APPARENT ABSENCE IN INTERMEDIATE SPECIES." J MOL BIOL **339**(3): 495-504.

JACKSON, R. J. "THE CURRENT STATUS OF VERTEBRATE CELLULAR MRNA IRESS." COLD SPRING HARB PERSPECT BIOL **5**(2).

JACKSON, R. J., C. U. HELLEN, ET AL. "THE MECHANISM OF EUKARYOTIC TRANSLATION INITIATION AND PRINCIPLES OF ITS REGULATION." NAT REV MOL CELL BIOL **11**(2): 113-127.

JANZEN, D. M., L. FROLOVA, ET AL. (2002). "INHIBITION OF TRANSLATION TERMINATION MEDIATED BY AN INTERACTION OF EUKARYOTIC RELEASE FACTOR 1 WITH A NASCENT PEPTIDYL-TRNA." MOL CELL BIOL **22**(24): 8562-8570.

JENNER, L., ET AL., STRUCTURAL REARRANGEMENTS OF THE RIBOSOME AT THE TRNA PROOFREADING STEP. NAT STRUCT MOL BIOL. 17(9): P. 1072-8.

KIM, D. F. AND R. GREEN (1999). "BASE-PAIRING BETWEEN 23S RRNA AND TRNA IN THE RIBOSOMAL A SITE." MOL CELL **4**(5): 859-864.

KLINGE, S., F. VOIGTS-HOFFMANN, ET AL. "CRYSTAL STRUCTURE OF THE EUKARYOTIC 60S RIBOSOMAL SUBUNIT IN COMPLEX WITH INITIATION FACTOR 6." SCIENCE **334**(6058): 941-948.

KHOSHNEVIS, S. ET AL. THE IRON–SULPHUR PROTEIN RNASE L INHIBITOR FUNCTIONS IN TRANSLATION TERMINATION. EMBO REP.11, 214–219 (2010).

KOSOLAPOV, A. AND C. DEUTSCH, TERTIARY INTERACTIONS WITHIN THE RIBOSOMAL EXIT TUNNEL. NAT STRUCT MOL BIOL, 2009. 16(4): P. 405-11.

KOLOSOV, P., L. FROLOVA, ET AL. (2005). "INVARIANT AMINO ACIDS ESSENTIAL FOR DECODING FUNCTION OF POLYPEPTIDE RELEASE FACTOR ERF1." NUCLEIC ACIDS RES **33**(19): 6418-6425.

LANG, K. J., A. KAPPEL, ET AL. (2002). "HYPOXIA-INDUCIBLE FACTOR-1ALPHA MRNA CONTAINS AN INTERNAL RIBOSOME ENTRY SITE THAT ALLOWS EFFICIENT TRANSLATION DURING NORMOXIA AND HYPOXIA." MOL BIOL CELL **13**(5): 1792-1801.

LIAO, P. Y., Y. S. CHOI, ET AL. "THE MANY PATHS TO FRAMESHIFTING: KINETIC MODELLING AND ANALYSIS OF THE EFFECTS OF DIFFERENT ELONGATION STEPS ON PROGRAMMED -1

RIBOSOMAL FRAMESHIFTING." NUCLEIC ACIDS RES **39**(1): 300-312.

LIU, B.S., X.Y. LIU, AND C. QIAN, [AN EFFICIENT TOOL FOR THE CONSTRUCTION OF MULTIPLE-CISTRONIC VECTORS: FMDV 2A]. SHENG WU GONG CHENG XUE BAO, 2007. 23(5): P. 765-9.

LUKE, G. A., P. DE FELIPE, ET AL. (2008). "OCCURRENCE, FUNCTION AND EVOLUTIONARY ORIGINS OF '2A-LIKE' SEQUENCES IN VIRUS GENOMES." J GEN VIROL **89**(PT 4): 1036-1042.

MILLIGAN, R. A. AND P. N. UNWIN (1986). "LOCATION OF EXIT CHANNEL FOR NASCENT PROTEIN IN 80S RIBOSOME." NATURE **319**(6055): 693-695.

MITRA, K., C. SCHAFFITZEL, ET AL. (2006). "ELONGATION ARREST BY SECM VIA A CASCADE OF RIBOSOMAL RNA REARRANGEMENTS." MOL CELL **22**(4): 533-543.

MOORE, M. J. (2005). "FROM BIRTH TO DEATH: THE COMPLEX LIVES OF EUKARYOTIC MRNAS." SCIENCE **309**(5740): 1514-1518.

MOUZAKIS, K. D., A. L. LANG, ET AL. "HIV-1 FRAMESHIFT EFFICIENCY IS PRIMARILY DETERMINED BY THE STABILITY OF BASE PAIRS POSITIONED AT THE MRNA ENTRANCE CHANNEL OF THE RIBOSOME." NUCLEIC ACIDS RES **41**(3): 1901-1913.

MUHLRAD, D., R. HUNTER, AND R. PARKER, A RAPID METHOD FOR LOCALIZED MUTAGENESIS OF YEAST GENES. YEAST, 1992. 8(2): P. 79-82.

MUNOZ, V. AND L. SERRANO, ELUCIDATING THE FOLDING PROBLEM OF HELICAL PEPTIDES USING EMPIRICAL PARAMETERS. III. TEMPERATURE AND PH DEPENDENCE. J MOL BIOL, 1995. 245(3): P. 297-308.

MUTO, H., H. NAKATOGAWA, ET AL. (2006). "GENETICALLY ENCODED BUT NONPOLYPEPTIDE PROLYL-TRNA FUNCTIONS IN THE A SITE FOR SECM-MEDIATED RIBOSOMAL STALL." MOL CELL **22**(4): 545-552.

NAKATOGAWA, H. AND K. ITO (2001). "SECRETION MONITOR, SECM, UNDERGOES SELF-TRANSLATION ARREST IN THE CYTOSOL." MOL CELL **7**(1): 185-192.

NAKATOGAWA, H. AND K. ITO (2002). "THE RIBOSOMAL EXIT TUNNEL FUNCTIONS AS A DISCRIMINATING GATE." CELL **108**(5): 629-636.

NAPTHINE, S., C. YEK, ET AL. "CHARACTERIZATION OF THE STOP CODON READTHROUGH SIGNAL OF COLORADO TICK FEVER VIRUS SEGMENT 9 RNA." RNA **18**(2): 241-252.

NAMY, O., ET AL., A MECHANICAL EXPLANATION OF RNA PSEUDOKNOT FUNCTION IN PROGRAMMED RIBOSOMAL FRAMESHIFTING. NATURE, 2006. 441(7090): P. 244-7.

NISSEN, P., J. HANSEN, ET AL. (2000). "THE STRUCTURAL BASIS OF RIBOSOME ACTIVITY IN PEPTIDE

BOND SYNTHESIS." SCIENCE **289**(5481): 920-930.

NOLLER, H. F. (1993). "PEPTIDYL TRANSFERASE: PROTEIN, RIBONUCLEOPROTEIN, OR RNA?" J BACTERIOL **175**(17): 5297-5300.

OGLE, J. M. AND V. RAMAKRISHNAN (2005). "STRUCTURAL INSIGHTS INTO TRANSLATIONAL FIDELITY." ANNU REV BIOCHEM **74**: 129-177.

ODON, V., ET AL., APE-TYPE NON-LTR RETROTRANSPOSONS OF MULTICELLULAR ORGANISMS ENCODE VIRUS-LIKE 2A OLIGOPEPTIDE SEQUENCES, WHICH MEDIATE TRANSLATIONAL RECODING DURING PROTEIN SYNTHESIS. MOL BIOL EVOL. 30(8): P. 1955-65.

OGLE, J.M., ET AL., RECOGNITION OF COGNATE TRANSFER RNA BY THE 30S RIBOSOMAL SUBUNIT. SCIENCE, 2001. 292(5518): P. 897-902.

PACE, C.N. AND J.M. SCHOLTZ, A HELIX PROPENSITY SCALE BASED ON EXPERIMENTAL STUDIES OF PEPTIDES AND PROTEINS. BIOPHYS J, 1998. 75(1): P. 422-7.

PISAREVA, V. P., A. V. PISAREV, ET AL. (2006). "KINETIC ANALYSIS OF INTERACTION OF EUKARYOTIC RELEASE FACTOR 3 WITH GUANINE NUCLEOTIDES." J BIOL CHEM **281**(52): 40224-40235.

PISAREV AV, ET AL. (2010) THE ROLE OF ABCE1 IN EUKARYOTIC POSTTERMINATION RIBOSOMAL RECYCLING. MOL CELL 37:196–210.

PLANT, E. P., K. L. JACOBS, ET AL. (2003). "THE 9-A SOLUTION: HOW MRNA PSEUDOKNOTS PROMOTE EFFICIENT PROGRAMMED -1 RIBOSOMAL FRAMESHIFTING." RNA **9**(2): 168-174.

POLACEK, N. AND A. S. MANKIN (2005). "THE RIBOSOMAL PEPTIDYL TRANSFERASE CENTER: STRUCTURE, FUNCTION, EVOLUTION, INHIBITION." CRIT REV BIOCHEM MOL BIOL **40**(5): 285-311.

RABL, J., M. LEIBUNDGUT, ET AL. "CRYSTAL STRUCTURE OF THE EUKARYOTIC 40S RIBOSOMAL SUBUNIT IN COMPLEX WITH INITIATION FACTOR 1." SCIENCE **331**(6018): 730-736.

REBORA, K., ET AL., YEAST AMP PATHWAY GENES RESPOND TO ADENINE THROUGH REGULATED SYNTHESIS OF A METABOLIC INTERMEDIATE. MOL CELL BIOL, 2001. 21(23): P. 7901-12.

REDONDO, N., M. A. SANZ, ET AL. "TRANSLATION DIRECTED BY HEPATITIS A VIRUS IRES IN THE ABSENCE OF ACTIVE EIF4F COMPLEX AND EIF2." PLOS ONE **7**(12): E52065.

RODNINA, M. V., K. B. GROMADSKI, ET AL. (2005). "RECOGNITION AND SELECTION OF TRNA IN TRANSLATION." FEBS LETT **579**(4): 938-942.

RODNINA, M. V. AND W. WINTERMEYER (2009). "RECENT MECHANISTIC INSIGHTS INTO EUKARYOTIC RIBOSOMES." CURR OPIN CELL BIOL **21**(3): 435-443.

ROHL, C.A., A. CHAKRABARTTY, AND R.L. BALDWIN, HELIX PROPAGATION AND N-CAP PROPENSITIES OF THE AMINO ACIDS MEASURED IN ALANINE-BASED PEPTIDES IN 40 VOLUME PERCENT TRIFLUOROETHANOL. PROTEIN SCI, 1996. 5(12): P. 2623-37.

RUSSELL, J. AND J. C. ZOMERDIJK (2006). "THE RNA POLYMERASE I TRANSCRIPTION MACHINERY." BIOCHEM SOC SYMP(73): 203-216.

RYAN, M. D. AND J. DREW (1994). "FOOT-AND-MOUTH DISEASE VIRUS 2A OLIGOPEPTIDE MEDIATED CLEAVAGE OF AN ARTIFICIAL POLYPROTEIN." EMBO J **13**(4): 928-933.

RYAN, M. D., A. M. KING, ET AL. (1991). "CLEAVAGE OF FOOT-AND-MOUTH DISEASE VIRUS POLYPROTEIN IS MEDIATED BY RESIDUES LOCATED WITHIN A 19 AMINO ACID SEQUENCE." J GEN VIROL **72 (PT 11)**: 2727-2732.

Ryan, M. D., Donnelly, M. L. L., Lewis, A., Mehrotra, A. P., Wilkie, J. & Gani, D. (1999). A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. Bioorg Chem 27, 55–79.

SEIT-NEBI, A., L. FROLOVA, AND L. KISSELEV, CONVERSION OF OMNIPOTENT TRANSLATION TERMINATION FACTOR ERF1 INTO CILIATE-LIKE UGA-ONLY UNIPOTENT ERF1. EMBO REP, 2002. 3(9): P. 881-6.

SCHMEING, T. M. AND V. RAMAKRISHNAN (2009). "WHAT RECENT RIBOSOME STRUCTURES HAVE REVEALED ABOUT THE MECHANISM OF TRANSLATION." NATURE **461**(7268): 1234-1242.

SEIDELT, B., C. A. INNIS, ET AL. (2009). "STRUCTURAL INSIGHT INTO NASCENT POLYPEPTIDE CHAIN-MEDIATED TRANSLATIONAL STALLING." SCIENCE **326**(5958): 1412-1415.

SIMONOVIC, M. AND T. A. STEITZ (2009). "A STRUCTURAL VIEW ON THE MECHANISM OF THE RIBOSOME-CATALYZED PEPTIDE BOND FORMATION." BIOCHIM BIOPHYS ACTA **1789**(9-10): 612-623.

Sikorski and Heiter, Genetics 122:19-27, 1989, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisia*e*

SHARMA, P., ET AL., 2A PEPTIDES PROVIDE DISTINCT SOLUTIONS TO DRIVING STOP-CARRY ON TRANSLATIONAL RECODING. NUCLEIC ACIDS RES. 40(7): P. 3143-51.

SIMS, R. J., 3RD, S. S. MANDAL, ET AL. (2004). "RECENT HIGHLIGHTS OF RNA-POLYMERASE-II-MEDIATED TRANSCRIPTION." CURR OPIN CELL BIOL **16**(3): 263-271.

STONELEY, M., S. A. CHAPPELL, ET AL. (2000). "C-MYC PROTEIN SYNTHESIS IS INITIATED FROM THE INTERNAL RIBOSOME ENTRY SEGMENT DURING APOPTOSIS." MOL CELL BIOL **20**(4): 1162-1169.

SZYMCZAK, A.L. AND D.A. VIGNALI, DEVELOPMENT OF 2A PEPTIDE-BASED STRATEGIES IN THE DESIGN OF MULTICISTRONIC VECTORS. EXPERT OPIN BIOL THER, 2005. 5(5): P. 627-38.

TAYLOR, D. J., J. NILSSON, ET AL. (2007). "STRUCTURES OF MODIFIED EEF2 80S RIBOSOME COMPLEXES REVEAL THE ROLE OF GTP HYDROLYSIS IN TRANSLOCATION." EMBO J **26**(9): 2421-2431.

VARSHAVSKY, A., THE N-END RULE PATHWAY OF PROTEIN DEGRADATION. GENES CELLS, 1997. 2(1): P. 13-28.

VOORHEES, R.M. AND V. RAMAKRISHNAN, STRUCTURAL BASIS OF THE TRANSLATIONAL ELONGATION CYCLE. ANNU REV BIOCHEM. 82: P. 203-36.

WANG, Z. AND M.S. SACHS, RIBOSOME STALLING IS RESPONSIBLE FOR ARGININE-SPECIFIC TRANSLATIONAL ATTENUATION IN NEUROSPORA CRASSA. MOL CELL BIOL, 1997. 17(9): P. 4904-13.

WILLIAMS, R.W., ET AL., SECONDARY STRUCTURE PREDICTIONS AND MEDIUM RANGE INTERACTIONS. BIOCHIM BIOPHYS ACTA, 1987. 916(2): P. 200-4.

WOOLHEAD, C. A., A. E. JOHNSON, ET AL. (2006). "TRANSLATION ARREST REQUIRES TWO-WAY COMMUNICATION BETWEEN A NASCENT POLYPEPTIDE AND THE RIBOSOME." MOL CELL **22**(5): 587-598.

YAN, F., V. A. DORONINA, ET AL. "ORCHESTRATING RIBOSOMAL ACTIVITY FROM INSIDE: EFFECTS OF THE NASCENT CHAIN ON THE PEPTIDYLTRANSFERASE CENTRE." BIOCHEM SOC TRANS **38**(6): 1576-1580.

YAP, M. N. AND H. D. BERNSTEIN (2009). "THE PLASTICITY OF A TRANSLATION ARREST MOTIF YIELDS INSIGHTS INTO NASCENT POLYPEPTIDE RECOGNITION INSIDE THE RIBOSOME TUNNEL." MOL CELL **34**(2): 201-211.

YONATH, A., K. R. LEONARD, ET AL. (1987). "A TUNNEL IN THE LARGE RIBOSOMAL SUBUNIT REVEALED BY THREE-DIMENSIONAL IMAGE RECONSTRUCTION." SCIENCE **236**(4803): 813-816.

YOSHINAKA, Y., I. KATOH, ET AL. (1985). "MURINE LEUKEMIA VIRUS PROTEASE IS ENCODED BY THE GAG-POL GENE AND IS SYNTHESIZED THROUGH SUPPRESSION OF AN AMBER TERMINATION CODON." PROC NATL ACAD SCI U S A **82**(6): 1618-1622.

YOSHINAKA, Y., I. KATOH, ET AL. (1985). "TRANSLATIONAL READTHROUGH OF AN AMBER TERMINATION CODON DURING SYNTHESIS OF FELINE LEUKEMIA VIRUS PROTEASE." J VIROL **55**(3): 870-873.

YU, Y., I. S. ABAEVA, ET AL. "COMMON CONFORMATIONAL CHANGES INDUCED IN TYPE 2 PICORNAVIRUS IRESS BY COGNATE TRANS-ACTING FACTORS." NUCLEIC ACIDS RES **39**(11): 4851-4865.

ZIV, G., G. HARAN, ET AL. (2005). "RIBOSOME EXIT TUNNEL CAN ENTROPICALLY STABILIZE ALPHA-HELICES." PROC NATL ACAD SCI U S A **102**(52): 18956-18961.